

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 03:15:53 ; Search time 8364 Seconds
(without alignments)
11136.320 Million cell updates/sec

Title: US-10-027-880-1

Perfect score: 2149

Sequence: 1 gtcgaattgcatattgt.....actaatgctaattaatcag 2149

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2149	100.0	2149	6	AX063409	AX063409 Sequence
2	2149	100.0	4280	6	AX063410	AX063410 Sequence
3	2149	100.0	4309	6	AX063412	AX063412 Sequence
4	2149	100.0	4413	6	AX063411	AX063411 Sequence
5	2149	100.0	4413	8	AF242314	AF242314 Arabidops
6	2149	100.0	82454	8	AC012193	AC012193 Arabidops
7	1733	80.6	1741	8	AF242313	AF242313 Arabidops
8	251.8	11.7	317	8	ATH533603	ATH533603 Arabidops
9	132	6.1	295	11	AL773135	AL773135 Arabidops
10	93.8	4.4	43982	8	AC007289	AC007289 Arabidops
11	76.2	3.5	343050	3	PFA929353	AL929353 Plasmodiu
12	71	3.3	175842	9	AL731547	AL731547 Human DNA
13	70.4	3.3	14867	3	AE001398	AE001398 Plasmodiu
14	70.4	3.3	347050	3	PFA929351	AL929351 Plasmodiu
15	69.2	3.2	104386	8	ATT32A17	AL161813 Arabidops
16	69.2	3.2	170627	2	AC125567	AC125567 Rattus no
17	69.2	3.2	179771	8	ATCHRIV25	AL161513 Arabidops
18	69	3.2	175544	2	AC117342	AC117342 Rattus no
19	68.4	3.2	250707	3	AE014848	AE014848 Plasmodiu
20	66.8	3.1	129240	9	AC084128	AC084128 Homo sapi
21	66.2	3.1	70098	8	AB025605	AB025605 Arabidops
22	66.2	3.1	73663	8	AF002029	AF002029 Arabidops
23	66	3.1	181792	9	AC098822	AC098822 Homo sapi
24	65.6	3.1	154995	9	AC011979	AC011979 Homo sapi
25	65.2	3.0	249995	3	AE014840	AE014840 Plasmodiu
26	64.6	3.0	111861	9	AC063435	AC063435 Homo sapi
27	64.6	3.0	156060	2	AC004153	AC004153 Plasmodiu
28	64.4	3.0	20529	2	AC116922	AC116922 Dictyoste
29	64.2	3.0	152824	2	EX663616	EX663616 Danio rer
30	64	3.0	144076	5	AL928677	AL928677 Zebrafish
31	63.8	3.0	183584	9	AC012492	AC012492 Homo sapi
32	63.8	3.0	189402	9	AC083829	AC083829 Homo sapi
33	63.6	3.0	7008	6	AX251286	AX251286 Sequence
34	63.4	3.0	526	9	HSFRAN1	AF053528 Homo sapi
35	63.4	3.0	110000	2	PFMAL1P1_07	Continuation (8 of
36	63.4	3.0	160661	2	EX548246	Continuation (8 of
37	63.2	2.9	148337	2	AC021509	EX548246 Danio rer
38	63.2	2.9	171255	9	AL691447	AC021509 Homo sapi
39	62.8	2.9	125623	3	AC115599	AL691447 Human DNA
40	62.6	2.9	331039	3	AC116988	AC115599 Dictyoste
41	62.4	2.9	110000	2	PFMAL13_00	AC116988 Dictyoste
42	62.4	2.9	110000	2	PFMAL13_23	AL844509 Plasmodiu
43	62.4	2.9	154890	2	EX548249	Continuation (24 o
44	62.2	2.9	1141	6	AX083744	EX548249 Danio rer
45	61.8	2.9	139111	9	AL158048	AX083744 Sequence
						AL158048 Human DNA

ALIGNMENTS

RESULT 1	AX063409	Sequence 1 from Patent WO0100833.	2149 bp	DNA	linear	PAT 24-JAN-2001
AX063409	LOCUS	AX063409				
DEFINITION	AX063409	Sequence 1 from Patent WO0100833.				
ACCESSION	AX063409.1	GI:12541197				
VERSION	AX063409.1	GI:12541197				
KEYWORDS	Arabidopsis thaliana (thale cress)					
SOURCE	Arabidopsis thaliana					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi					
REFERENCE	Hoffmann,B., Mollner,P. and Pelletier,G.					
AUTHORS	Promoter expressed specifically in the cells of plant roots,					
TITLE						

recombinant vectors and host cells comprising same and transgenic

plants obtained

JOURNAL Patent: WO 0100833-A 1 04-JAN-2001;

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)

FEATURES Location/Qualifiers

1..2149

/organism="Arabidopsis thaliana"

/mol_type="unassigned DNA"

/db_xref="taxon:3702"

ORIGIN

Query Match 100.0%; Score 2149; DB 6; Length 2149;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GTGCAATTTGATATATTTGATGCAATCTGAAAGAAATAAGTGGGATATATAACAACCG	60
Db	1	GTGCAATTTGATATATTTGATGCAATCTGAAAGAAATAAGTGGGATATATAACAACCG	60
QY	61	GGCAAGTACAAAGTTCACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT	120
Db	61	GGCAAGTACAAAGTTCACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT	120
QY	121	AATTCCTGAATCTTCAATTCCTGAATGATATTTACATTTTATCAAAAAAAGTACA	180
Db	121	AATTCCTGAATCTTCAATTCCTGAATGATATTTACATTTTATCAAAAAAAGTACA	180
QY	181	AGTTCTACCAAGCACAGGAGTTAAACAACCTTGTTGTCCTCAATTCCTCAATTCCTCA	240
Db	181	AGTTCTACCAAGCACAGGAGTTAAACAACCTTGTTGTCCTCAATTCCTCAATTCCTCA	240
QY	241	TCTATGATTTCCCTTTCTTCCGATATATCTGATTTGATGATGATGATGATGATGATG	300
Db	241	TCTATGATTTCCCTTTCTTCCGATATATCTGATTTGATGATGATGATGATGATGATG	300
QY	301	TCATTAATCCCACTCTATACATCAGTATCTCAAGTCCAAATAACAATATCCATAAGAA	360
Db	301	TCATTAATCCCACTCTATACATCAGTATCTCAAGTCCAAATAACAATATCCATAAGAA	360
QY	361	GTGGTATATTGTGAAAAAAGGAGTGTATCTGGTATATACAAATACACGGTC	420
Db	361	GTGGTATATTGTGAAAAAAGGAGTGTATCTGGTATATACAAATACACGGTC	420
QY	421	TGCAATTCCTCAACATTTCTAGGAGAAATCGAGTGTCTCTTTGGTTTATTTATTT	480
Db	421	TGCAATTCCTCAACATTTCTAGGAGAAATCGAGTGTCTCTTTGGTTTATTTATTTAT	480
QY	481	CTTAATAACATATCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTTCAATGTGC	540
Db	481	CTTAATAACATATCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTTCAATGTGC	540
QY	541	TAAATTTCTATATCAATAATCGTAAGAAAAATTCGTCGAAGCCACAGGACATGCTATG	600
Db	541	TAAATTTCTATATCAATAATCGTAAGAAAAATTCGTCGAAGCCACAGGACATGCTATG	600
QY	601	GGCAGTATGTTACCTTTTAAACCATCAAAATATATTAAAGAAAAAGAACTTCTTAA	660
Db	601	GGCAGTATGTTACCTTTTAAACCATCAAAATATATTAAAGAAAAAGAACTTCTTAA	660
QY	661	AGAACATTTTAAAGTGGATAAAAAAGATAAAGAGTGGCAGAGAAAAACGATG	720
Db	661	AGAACATTTTAAAGTGGATAAAAAAGATAAAGAGTGGCAGAGAAAAACGATG	720
QY	721	CCGGACTCGTAACAGGAGCTCCGACCATCTCGGAGAGCGGACGCTGATGATTT	780
Db	721	CCGGACTCGTAACAGGAGCTCCGACCATCTCGGAGAGCGGACGCTGATGATTT	780
QY	781	TTTTCTTTTCTTTTCTTAAAGACGTTGTTTCGTGCTTACAGGGTCAAAACCATATCC	840
Db	781	TTTTCTTTTCTTTTCTTAAAGACGTTGTTTCGTGCTTACAGGGTCAAAACCATATCC	840
QY	841	AATTGTTCTGCTATTTATATATTAACCTTAAGATCCCTCTTGCTTTGCTTTTATTCGT	900
Db			

Db	841	AATTGTTCTGCTATTTATTTATATACTAAAGATCCCTCTCTGTGCTTTGTCTTTATTCGT	900
QY	901	GATATATAATCTAACTTAAATTTAGTTCTTAAATATATATATGCTTACTATGTTTCTACTG	960
Db	901	GATATATAATCTAACTTAAATTTAGTTCTTAAATATATATATGCTTACTATGTTTCTACTG	960
QY	961	ACCTGAGTCCCTAGTTAGCTATATGACATATGCTGAAATGACGCCCAAAATTTGAAGAG	1020
Db	961	ACCTGAGTCCCTAGTTAGCTATATGACATATGCTGAAATGACGCCCAAAATTTGAAGAG	1020
QY	1021	TTCTCTTCTGCACTAACTCTTATCTTACTCAITGAGCTATGTTTAAATATTAATGATTT	1080
Db	1021	TTCTCTTCTGCACTAACTCTTATCTTACTCAITGAGCTATGTTTAAATATTAATGATTT	1080
QY	1081	GGCAGTCTCGTATTAATATGCGAGTTGCACCTAGATATAAAACACATGATAGACATTTAG	1140
Db	1081	GGCAGTCTCGTATTAATATGCGAGTTGCACCTAGATATAAAACACATGATAGACATTTAG	1140
QY	1141	TTTAAACCTTGAATGTTATTTGAACTCTTTTGGATTACGTTGGATTGTTGATGATTTAA	1200
Db	1141	TTTAAACCTTGAATGTTATTTGAACTCTTTTGGATTACGTTGGATTGTTGATGATTTAA	1200
QY	1201	TTTGAAGATATTTATATATGAAGATGTTTATATATATAGAGTTTTATAGAGTAA	1260
Db	1201	TTTGAAGATATTTATATATGAAGATGTTTATATATATAGAGTTTTATAGAGTAA	1260
QY	1261	TATTGATGATGATGTTGCTCTTTTGTAGTTACTCTTTTGTGCGTAGTCTTCTCTCT	1320
Db	1261	TATTGATGATGATGTTGCTCTTTTGTAGTTACTCTTTTGTGCGTAGTCTTCTCTCTCT	1320
QY	1321	CATCTCTCTATGAAGAAAAATCCAAATAGTTTAAAGGAAATTTTGTGAATTCATAGTCT	1380
Db	1321	CATCTCTCTATGAAGAAAAATCCAAATAGTTTAAAGGAAATTTTGTGAATTCATAGTCT	1380
QY	1381	TTTTGTAACCAACAGTTCTATGATGATGATGATGATGATGATGATGATGATGATGATG	1440
Db	1381	TTTTGTAACCAACAGTTCTATGATGATGATGATGATGATGATGATGATGATGATGATG	1440
QY	1441	AGATCGTTTCTGTAATAATTTAGTAGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAA	1500
Db	1441	AGATCGTTTCTGTAATAATTTAGTAGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAA	1500
QY	1501	TCGAGCGTCTGTCATCCCAATAAGTTGTTGATTCGAGCTTCCAAAGATTAATATCTT	1560
Db	1501	TCGAGCGTCTGTCATCCCAATAAGTTGTTGATTCGAGCTTCCAAAGATTAATATCTT	1560
QY	1561	TTTTAGATGGGTGATGAGATTTCTAACTTCGATATAGGATGATGATGATGATGATGATG	1620
Db	1561	TTTTAGATGGGTGATGAGATTTCTAACTTCGATATAGGATGATGATGATGATGATGATG	1620
QY	1621	CATATAGTCTGTTTGTGAGCTCTGCTCTTTTGGAGACCAACCCCTTGTCTAATGTT	1680
Db	1621	CATATAGTCTGTTTGTGAGCTCTGCTCTTTTGGAGACCAACCCCTTGTCTAATGTT	1680
QY	1681	TTCTTGCACTTAGACAAATCCATAACGTTACGTTAGTTCGAGTTCGACCAAAATGGTC	1740
Db	1681	TTCTTGCACTTAGACAAATCCATAACGTTACGTTAGTTCGAGTTCGACCAAAATGGTC	1740
QY	1741	CAATATATTTTAAATTTGCGCACCAAAACAACTTTTACAAACAAATTTCAACAAACATGC	1800
Db	1741	CAATATATTTTAAATTTGCGCACCAAAACAACTTTTACAAACAAATTTCAACAAACATGC	1800
QY	1801	ATCGTTTCAAAATTTTATTTTCAATGCGCTTATTTGTTTCAATTTGTAATATTTCTGTTAA	1860
Db	1801	ATCGTTTCAAAATTTTATTTTCAATGCGCTTATTTGTTTCAATTTGTAATATTTCTGTTAA	1860
QY	1861	CTCAGTGAAGATTTTAAATTTTCAAGAGAAACATTTTGTATATTAATTAATTAATTTTA	1920
Db	1861	CTCAGTGAAGATTTTAAATTTTCAAGAGAAACATTTTGTATATTAATTAATTAATTTTA	1920
QY	1921	TGGAACCAACCGGTTAGCTGATGATTTTCAAGTATTTAGTTTGTGCTGTTTGTGAAATCAT	1980
Db	1921	TGGAACCAACCGGTTAGCTGATGATTTTCAAGTATTTAGTTTGTGCTGTTTGTGAAATCAT	1980


```
Db 1501 TCTAGCGTCTGTCAATCCCAAAATAAGTTGTTTGAATCGAGCTTCAAGATTATAATCTT 1560
Qy 1561 TTTTAGATGGGTCATGAAGATTCTAACTTCGTATACGAGTGATCCATATAATTTCTAA 1620
Db 1561 TTTTAGATGGGTCATGAAGATTCTAACTTCGTATACGAGTGATCCATATAATTTCTAA 1620
Qy 1621 CATATACGCTCTGTTTGTGTAGCTCTGCGTCTTTTGAGACACACCCCTTGTAAATGTT 1680
Db 1621 CATATACGCTCTGTTTGTGTAGCTCTGCGTCTTTTGAGACACACCCCTTGTAAATGTT 1680
Qy 1681 TTGTTGACCTTAGACAATCCATAATACGTTTACGTGAGTGAAGTTGCACCAAAATGGTC 1740
Db 1681 TTGTTGACCTTAGACAATCCATAATACGTTTACGTGAGTGAAGTTGCACCAAAATGGTC 1740
Qy 1741 CAAATATAATTTAAATTTGGCCCAAAACAACATTTTACAAAACAAATTCACAAACATGC 1800
Db 1741 CAAATATAATTTAAATTTGGCCCAAAACAACATTTTACAAAACAAATTCACAAACATGC 1800
Qy 1801 ATCGTTTCAAAATTTATTTTCAATGGCGTTATTTGTTTCAATTTGTAATATTTCTGTTAA 1860
Db 1801 ATCGTTTCAAAATTTATTTTCAATGGCGTTATTTGTTTCAATTTGTAATATTTCTGTTAA 1860
Qy 1861 CTCACGTACGAATTTTAAATTTTCAAGAAGAACATTTTGTATATAAAATAACATTTTA 1920
Db 1861 CTCACGTACGAATTTTAAATTTTCAAGAAGAACATTTTGTATATAAAATAACATTTTA 1920
Qy 1921 TGGAAACACCGGTTAAGCTCGATGATTTGAGTTTATGTTTGTGTTTGTGAAATCAT 1980
Db 1921 TGGAAACACCGGTTAAGCTCGATGATTTGAGTTTATGTTTGTGTTTGTGAAATCAT 1980
Qy 1981 TAAGACCTACATTTGATCCCTCAATTAATTAATTAAGATTAAGATTAAGATTAAGATTA 2040
Db 1981 TAAGACCTACATTTGATCCCTCAATTAATTAATTAAGATTAAGATTAAGATTAAGATTA 2040
Qy 2041 GTTCACCAAAAGACGTCCTTATGAGTATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2100
Db 2041 GTTCACCAAAAGACGTCCTTATGAGTATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2100
Qy 2101 TAAGACGCTTATATTCAACATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 2149
Db 2101 TAAGACGCTTATATTCAACATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 2149

RESULT 3
AX063412
LOCUS Sequence 4 from Patent WO0100833.
DEFINITION AX063412
ACCESSION AX063412
VERSION AX063412.1 GI:12541200
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Hoffmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots, recombinant vectors and host cells comprising same and transgenic plants obtained
JOURNAL Patent: WO 0100833-A 4 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES
Location/Qualifiers
source 1..4309
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Insert du vecteur pBin19"

ORIGIN
Query Match 100.0%; Score 2149; DB 6; Length 4309;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGAATTTGTGATATTGTAAGCAATCTGAAAGAAATAGTGGGATATATAAACACCG 60
```

```
Db 30 GTCGAATTTGTGATATTGTAAGCAATCTGAAAGAAATAGTGGGATATATAAACACCG 89
Qy 61 GCGAAAGTACAAGTTCTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 120
Db 90 GCGAAAGTACAAGTTCTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 149
Qy 121 AATTCCTGAATCTTTTCATTTCTGAATTTGATTTTATCAATTTTATCAAAAATAAGATACA 180
Db 150 AATTCCTGAATCTTTTCATTTCTGAATTTGATTTTATCAATTTTATCAAAAATAAGATACA 209
Qy 181 AGTTCTACCAAGACACAGGAGTTTAAACAACCTTGTGTGTCAAATTCGTAATTTAAAGCCTAA 240
Db 210 AGTTCTACCAAGACACAGGAGTTTAAACAACCTTGTGTGTCAAATTCGTAATTTAAAGCCTAA 269
Qy 241 TCTATGATTTCCCTTTTCTTCAGATATATACATGATTTGATATGACACCATTTGTTG 300
Db 270 TCTATGATTTCCCTTTTCTTCAGATATATACATGATTTGATATGACACCATTTGTTG 329
Qy 301 TCATTAACCTTCCACCTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
Db 330 TCATTAACCTTCCACCTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 389
Qy 361 GTGGTATATTGTGAAAAAATAAGTGTGTATCTGGTATATACATTAACAATCCACGGTC 420
Db 390 GTGGTATATTGTGAAAAAATAAGTGTGTATCTGGTATATACATTAACAATCCACGGTC 449
Qy 421 TCGAATTCCTCAACAATTTCTAGGAGAAATCGAGTGTCTCTTGGTTTATTTATT 480
Db 450 TCGAATTCCTCAACAATTTCTAGGAGAAATCGAGTGTCTCTTGGTTTATTTATT 509
Qy 481 CTTAATAACATCTCTATATTTTAAACACCTTCGATGTCTCGCTTAAATTTTGAATGTGCC 540
Db 510 CTTAATAACATCTCTATATTTTAAACACCTTCGATGTCTCGCTTAAATTTTGAATGTGCC 569
Qy 541 TAAATTTCTCTAATCAATAAATCGTAAAGAAAAATTCGTCGAAGCCACAGGACATGCAATAG 600
Db 570 TAAATTTCTCTAATCAATAAATCGTAAAGAAAAATTCGTCGAAGCCACAGGACATGCAATAG 629
Qy 601 GGCACGTAGTTAGCTTTTAAACCATTAATAATATATTAAGAAAGGAACTTCCCTAAA 660
Db 630 GGCACGTAGTTAGCTTTTAAACCATTAATAATATATTAAGAAAGGAACTTCCCTAAA 689
Qy 661 AGAACATTTAATAAGTGGATATAAAGATAGAGGTAGGAGAGAAAGAAAGCTATGG 720
Db 690 AGAACATTTAATAAGTGGATATAAAGATAGAGGTAGGAGAGAAAGAAAGCTATGG 749
Qy 721 CCGCGACTCGTAAACAGGGACGTCCTCGACCACTCGCGAGACGGCGAGACGCTGATTT 780
Db 750 CCGCGACTCGTAAACAGGGACGTCCTCGACCACTCGCGAGACGGCGAGACGCTGATTT 809
Qy 781 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTGCTGTACAGGGTCACAAACCATATCC 840
Db 810 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTGCTGTACAGGGTCACAAACCATATCC 869
Qy 841 AATTTGTTCTGCTATTATTAATACTAAAGATCCCTCTTGTCTTGTCTTTTGTCTTATCGT 900
Db 870 AATTTGTTCTGCTATTATTAATACTAAAGATCCCTCTTGTCTTGTCTTTTGTCTTATCGT 929
Qy 901 GATATATAATCTAACTTAAATTTAGTTCTTAAATAATATATATGTCCTACCTATGTTTCTACTIG 960
Db 930 GATATATAATCTAACTTAAATTTAGTTCTTAAATAATATATATGTCCTACCTATGTTTCTACTIG 989
Qy 961 ACCTCAGTCCCTAGTTAGCTATATGACATATGAAAAATGACCCCAAAATTTTGAAGAG 1020
Db 990 ACCTCAGTCCCTAGTTAGCTATATGACATATGAAAAATGACCCCAAAATTTTGAAGAG 1049
Qy 1021 TTCTCTTCTCTGCAACTAACTCTTATCTTACTCTAGCTATGTTTAAATTTGAATGTT 1080
Db 1050 TTCTCTTCTCTGCAACTAACTCTTATCTTACTCTAGCTATGTTTAAATTTGAATGTT 1109
Qy 1081 GGCACTCTCGTATTAAATATGCCAGTTGCACCTAGATAAAAAACATGATAGACATTTAG 1140
```


[illegible]

QY 61 GCGAAGTACGAAGTCTACCTTTTGGCATGGAACCATGTTTTAGGATTTACTTGT 120
Db 196 GCGAAGTACGAAGTCTACCTTTTGGCATGGAACCATGTTTTAGGATTTACTTGT 255
QY 121 AATTCTCGAATCTTTCATTTCTGGAATGATATTTACATTTTATCAAAAAAAGTACA 180
Db 256 AATTCTCGAATCTTTCATTTCTGGAATGATATTTACATTTTATCAAAAAAAGTACA 315
QY 181 AGTTCTACCAAGCACAGAGTTAAACAACTTGCTGTCMAATGCTAATTTAAAGCCTAA 240
Db 316 AGTTCTACCAAGCACAGAGTTAAACAACTTGCTGTCMAATGCTAATTTAAAGCCTAA 375
QY 241 TCTTATGATTTCCCTTTTCTCAAGATATATCTGATATGATATGCAACCCATTTGTTG 300
Db 376 TCTTATGATTTCCCTTTTCTCAAGATATATCTGATATGATATGCAACCCATTTGTTG 435
QY 301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAGTCGAATACATATCCATAAGAA 360
Db 436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAGTCGAATACATATCCATAAGAA 495
QY 361 GTGATATATTGTGAAAAAAGAAAAAGTGGTATATCTGGTATATACAAATACCAACGTC 420
Db 496 GTGATATATTGTGAAAAAAGAAAAAGTGGTATATCTGGTATATACAAATACCAACGTC 555
QY 421 TCGAATTGCTCAACAATTTCTAGAGAAAATGGAAGTGTCTCTTTGGTTTATTTATT 480
Db 556 TCGAATTGCTCAACAATTTCTAGAGAAAATGGAAGTGTCTCTTTGGTTTATTTATT 615
QY 481 CTTAATAACATACATATATTTTAAACACTTCGATCTCGCTTAAATTTTCGAATGTC 540
Db 616 CTTAATAACATACATATATTTTAAACACTTCGATCTCGCTTAAATTTTCGAATGTC 675
QY 541 TAAATTTCTCTAATCATATAATCGTAAAGAAAATCGTGAAGCCACAGGACATGCATAG 600
Db 676 TAAATTTCTCTAATCATATAATCGTAAAGAAAATCGTGAAGCCACAGGACATGCATAG 735
QY 601 GGCAAGTAGTTTACCTTTAAACACATATTAATTAAGAAAAGAAAATTTCTTAA 660
Db 736 GGCAAGTAGTTTACCTTTAAACACATATTAATTAAGAAAAGAAAATTTCTTAA 795
QY 661 AGAACATTTAATAAGTGGATTAAGAAAAGTAAAGTAGGAGGAGAGAAAAGTATGG 720
Db 796 AGAACATTTAATAAGTGGATTAAGAAAAGTAAAGTAGGAGGAGAGAAAAGTATGG 855
QY 721 CCGGACTCGTAAACAAGGAGCTCCGACCACTCGGAGACGGGAGACGCTGATGAT 780
Db 856 CCGGACTCGTAAACAAGGAGCTCCGACCACTCGGAGACGGGAGACGCTGATGAT 915
QY 781 TTTTCTTTTCTTTTCTTAAGAACGTTGTTTGGTGTCTACAAGGGTCAAAACCATATCC 840
Db 916 TTTTCTTTTCTTTTCTTAAGAACGTTGTTTGGTGTCTACAAGGGTCAAAACCATATCC 975
QY 841 AATTGTTCTGCTATTATATATAACTAAAGATCCCTCTGCTGCTTGTGCTTTATTCGT 900
Db 976 AATTGTTCTGCTATTATATATAACTAAAGATCCCTCTGCTGCTTGTGCTTTATTCGT 1035
QY 901 GATATATAATCTAACTTAAATTTAGTTCTTAAATATATATGTCCTACCTATGTTTCTACTG 960
Db 1036 GATATATAATCTAACTTAAATTTAGTTCTTAAATATATATGTCCTACCTATGTTTCTACTG 1095
QY 961 ACCTCAGTCCCTAGTTAGCTATATGGAATATGGAATAATGACCCCAAAATTTGAAGAG 1020
Db 1096 ACCTCAGTCCCTAGTTAGCTATATGGAATATGGAATAATGACCCCAAAATTTGAAGAG 1155
QY 1021 TTCTCTCTCTGCAACTAACTCTATCTTACTGATGAGCTATGTTTAAATATTGAATGTT 1080
Db 1156 TTCTCTCTCTGCAACTAACTCTTATCTTACTGATGAGCTATGTTTAAATATTGAATGTT 1215
QY 1081 GGCACTCTGATTTAAATATGCAAGTTGCACTAGATAAAAAAATCATGATAGACATTTAG 1140
Db 1216 GGCACTCTGATTTAAATATGCAAGTTGCACTAGATAAAAAAATCATGATAGACATTTAG 1275

QY 1141 TTTAAAACCTGAAATGTTTATTTGAACCTTTTGGATTACGTGGATTTGTTGATGGATTAA 1200
Db 1276 TTTAAAACCTGAAATGTTTATTTGAACCTTTTGGATTACGTGGATTTGTTGATGGATTAA 1335
QY 1201 TTTTGAAGATATTTATATATTAATTAAGATGTTTATATATATAGAGTTTATATAGCAGAAA 1260
Db 1336 TTTTGAAGATATTTATATATTAATTAAGATGTTTATATATATAGAGTTTATATAGCAGAAA 1395
QY 1261 TATTGATCTAGATGTTGCTTTTGTAGTACTCTTTTGTGTCGTAGTCTCTCTCTCT 1320
Db 1396 TATTGATCTAGATGTTGCTTTTGTAGTACTCTTTTGTGTCGTAGTCTCTCTCTCTCT 1455
QY 1321 CATCTCTCTATGAAGAAAATCCAAATAGTTTAAAGAAAATTTTGTGTAATTCATAGTCT 1380
Db 1456 CATCTCTCTATGAAGAAAATCCAAATAGTTTAAAGAAAATTTTGTGTAATTCATAGTCT 1515
QY 1381 TTTTCGTAAACACAGTCTATGTAGCTATGCTATCATATTTCTCTTTGCAACAAACAAA 1440
Db 1516 TTTTCGTAAACACAGTCTATGTAGCTATGCTATCATATTTCTCTTTGCAACAAACAAA 1575
QY 1441 AAGATCGTTTGTGAAAATTTAGTAGGCACTTAAAGTCGTCTATTTGTTGTCCTCTCT 1500
Db 1576 AAGATCGTTTGTGAAAATTTAGTAGGCACTTAAAGTCGTCTATTTGTTGTCCTCTCT 1635
QY 1501 TCTAGCGTCTCTGTCATCCACAAATAAGTTGTTGATTCGAGCTTCCAAAGATTATAATCTT 1560
Db 1636 TCTAGCGTCTCTGTCATCCACAAATAAGTTGTTGATTCGAGCTTCCAAAGATTATAATCTT 1695
QY 1561 TTTTAGATGGGTGATGAAGATTTCTAATCTCGTATACGAGTATGTCATATAATTTCTTAA 1620
Db 1696 TTTTAGATGGGTGATGAAGATTTCTAATCTCGTATACGAGTATGTCATATAATTTCTTAA 1755
QY 1621 CATATAGCTCTGTTTGTGTTAGTCTCTGCTCTTTTGAGACCACTCCCTTCTGCTTAA 1680
Db 1756 CATATAGCTCTGTTTGTGTTAGTCTCTGCTCTTTTGAGACCACTCCCTTCTGCTTAA 1815
QY 1681 TTGTTGACCTTAGACAATCCATAATACGTTTACGTGAGTTCGAAGTTGCACCAAAATGCTC 1740
Db 1816 TTGTTGACCTTAGACAATCCATAATACGTTTACGTGAGTTCGAAGTTGCACCAAAATGCTC 1875
QY 1741 CAAATATATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTCACAAACATGC 1800
Db 1876 CAAATATATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTCACAAACATGC 1935
QY 1801 ATCTGTTTCAAAATTTATTTATTCATGCGGTTATTTGTTGATTTGTAATATTTCTGTTAA 1860
Db 1936 ATCTGTTTCAAAATTTATTTATTCATGCGGTTATTTGTTGATTTGTAATATTTCTGTTAA 1995
QY 1861 CTCACTGACGAATTTTAAATTTTCAAGAGAACAATTTTGTATATAAAATTAACATTTTA 1920
Db 1996 CTCACTGACGAATTTTAAATTTTCAAGAGAACAATTTTGTATATAAAATTAACATTTTA 2055
QY 1921 TGGNACCAACCGGTTAAGCTCGATGATTTGAGTTTATGTTTGTGCTGTTTGTGAAATCAT 1980
Db 2056 TGGNACCAACCGGTTAAGCTCGATGATTTGAGTTTATGTTTGTGCTGTTTGTGAAATCAT 2115
QY 1981 TAAAGACCTACATTTGATCCCTCATTTACTTTTAAATTTAGGAATCAACATGATGATTAA 2040
Db 2116 TAAAGACCTACATTTGATCCCTCATTTACTTTTAAATTTAGGAATCAACATGATGATTAA 2175
QY 2041 GTTTCACCAAGAGCTCTCTTATGCTTATTAAGAGTTCAGACGCAAGGATGATGATGATGAT 2100
Db 2176 GTTTCACCAAGAGCTCTCTTATGCTTATTAAGAGTTCAGACGCAAGGATGATGATGATGAT 2235
QY 2101 TAAAGAGCTCTTATTTTCAACACCTTCTCTTATTAAGAGTTCAGACGCAAGGATGATGATGAT 2149
Db 2236 TAAAGAGCTCTTATTTTCAACACCTTCTCTTATTAAGAGTTCAGACGCAAGGATGATGATGAT 2284

RESULT 6

AC012193/c

LOCUS

DEFINITION Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence, AC012193 82454 bp DNA linear PLN 19-JAN-2001

```

complete sequence.
AC012193
VERSION AC012193.6 GI:12323286
SOURCE HTG.
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 82454)
Lin.X. Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Niernan,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence
Unpublished
2 (bases 1 to 82454)
Lin.X. and Kaul,S.
Direct Submission
Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 82454)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280880.
Address all correspondence to:at@tigr.org

BAC clone T32E8 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/edb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1. 82454
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T32E8"
/complement(1..3320)
/gene="T32E8.1"
/notes="similar to ribonuclease II RNb family protein
GB:CA16367 [Schizosaccharomyces pombe]; Pfam HMM hit:
vacB and RNase II family 3'-5' exoribonucleases"
complement(join(<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..>3320))
/gene="T32E8.1"
complement(join(<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..3320))
/gene="T32E8.1"
/codon_start=1
/product="putative 3'-5' exoribonuclease, 3' partial;
3320-1"

FEATURES
source
1. 82454
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T32E8"
/complement(1..3320)
/gene="T32E8.1"
/notes="similar to ribonuclease II RNb family protein
GB:CA16367 [Schizosaccharomyces pombe]; Pfam HMM hit:
vacB and RNase II family 3'-5' exoribonucleases"
complement(join(<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..>3320))
/gene="T32E8.1"
complement(join(<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..3320))
/gene="T32E8.1"
/codon_start=1
/product="putative 3'-5' exoribonuclease, 3' partial;
3320-1"

protein_id="AAG51632.1"
/db_xref="GI:12323300"
/translation="MKASSEQSVIERIENGKKRNRPKQNRKSKQSSVPIEDAHVE
ESLDGRDSRSRAKOSTSSSKQORNTDELEARNANVAFNSPMPRAESGVPFRASAS
PLSSPEVSKQLLSRNNRNEAYCKIDGVTDLIRNGVQCSRAVEDGVVTKDPLSL
LEKGFALFRVNAHNRNEAYCKIDGVTDLIRNGVQCSRAVEDGVVTKDPLSL
WPKMGVFTESAAPKPGTSPKQKDKKARQKNGIDVVEGFEDGFSKNSKSVTKGAK
NKTWTFSSPDLSCIGSCFCEQKNGCAVQKLCGILSSFPKRPFGQWAVKSLVRD
SVGLLDVKGWHIHYKESDPKCKSPLSLSDDVYVQLMPADRPFKLIVPFLVPSGIR
ARLENDPNELEAELVAQIVDNGESPPFVAQITHLFGSGSELEPOINALYONSVD
SDFSPGTSLSRPRVWEVEEVRQRKRDLCVLITDSTATDLDALSVQSLPGGF
FRVGVHIADVSFVLPETALDTEARFSTSVIMQRKISMLPPLSENGWSPAGDR
LAFSLIILNDNREGDVIIRWIGRTIIRSCCKLSYDHAQDIIDGKSDVENGSPALHGSF
KWCDFTRSVKQLSESTTLROKRFNGALQNSKVFLLKDEHGVDFVFTVSRKGSN
FLVBERPMLANMTAAEVSQAYRASSLLRHPEBNTKRLKEPEGFCSEKGMDLDISS
GQIQSLEKITGNLKDSDSVFVILNNYAIKPMQLASYFCTGNLKDSEVGHVALAVP
LYTHFTSPLRYPDIVVHRALAALEAEELYSKQKQTAIDEGRCSTGIHFNKDAEIS
IRGKEALSVAALKGVSPSTEILSDVAAYCNERKLAARKVRDADCKLYTVPFLKQKEIF
PCEARVMNLGSRFMTVYISKL"
1. 189
/notes="overlap with BAC clone T5M16
[AC010704:103165..103353]"
/complement(8589..10716)
/gene="T32E8.2"
/notes="similar to AUX1 (regulator of root gravitropism,
putative permease) GB:CAA67308 [Arabidopsis thaliana]"
complement(join(<8589..8783,9000..9465,9555..9652,
9753..9936,10020..10132,10236..10421,10504..>10716))
/gene="T32E8.2"
complement(join(8589..8783,9000..9465,9555..9652,
9753..9936,10020..10132,10236..10421,10504..>10674))
/gene="T32E8.2"
/codon_start=1
/product="putative AUX1-like permease; 10674-8589"
/protein_id="AAG51630.1"
/db_xref="GI:12323298"
/translation="MAAEKIEIVVAGNYLEMEBEENISGNKKSSTKTLSNFWHGG
SVYDAMPSCANQVAQLVLTLPYSFQSGMGSLFLQFYGLMSQWTAYLIVLYVEY
RTKREKEDFRNHVQWPEVLIDGLKHWNLGLIFNCTFLFGSVQIACASNY
YINDKLDKRTWYIFGACCATVFIPEFNHYTWSFLGLAMTYTQWYLTIASLHHQ
AEDVSGDGTWVLYFTGATNLTLYFGGHAVTYEIMAMKPKQKFAVLLATLYLT
LTLPASAVYAFAGDKLLTHSNALSKLPKGFRTAVILMLHQFTFGFASPLFIV
WEKLGHTKFMKFKEMARLPVVPVFIALLIFPPFGPINSVAGSLVLSFTVYIIIPA
LAHMLTFAPAPSENAVERPVRVVGMMGTICINIFVVMVVFVVGFGFGWASMVNFV
RQIDTGLFTKCYQCPHPK"
12104..13574
/gene="T32E8.3"
/notes="similar to thaumatin-like protein GB:AA871214
[Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family"
join(<12104..12179,12505..12628,12704..>13574)
/gene="T32E8.3"
join(12104..12179,12505..12628,12704..13574)
/gene="T32E8.3"
/codon_start=1
/product="thaumatin-like protein; 12104-13574"
/protein_id="AAG51631.1"
/db_xref="GI:12323299"
/translation="MDRTIEEDINGPVSILIRPNISVASLTSRSNNITFDPKIFS
SKDANSFRFVKDLVTLSSHIIRVMTIRTPVNVSTGKWSARVFTIINSQDQTIW
PALTPGENFNGGFFELKPGQSIYFHAPVGSGRIGWRTGCKFDSTGTCTGCGST
LKCSASGKPPASIAEFTLAALDFYDLSVDGFLNPSVTMNGKNGSVAGCVADLRP
HCPQELAVKSNKGVIICRSACDFDRDEYCCRGVYGVNVVVCQPTYSKIPKACPTAY
SVAYDDPTSIMTCTASDYVISFCSEFRVLTFTLTSHDYFNDSQVSFLDHDFFYI
MIYFIHPLILISYAKVIQWQK"
14054..15303
/gene="T32E8.4"
/notes="similar to hypothetical protein GB:P34661
[Ceanorhabditis elegans]"
join(14054..14174,14251..14348,14779..14811,15170..15303)
/gene="T32E8.4"
join(14107..14174,14251..14348,14779..14811,15170..15252)
/gene="T32E8.4"
/codon_start=1

```


	FEATURES	Location/Qualifiers
	source	1..1741 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /map="NIAL-agg64" /ecotype="Wassilewskija"
	misc_feature	<1..>1741 /note="contains root-specific promoter"
	ORIGIN	
		Query Match 80.6%; Score 1733; DB 8; Length 1741; Best Local Similarity 99.7%; Pred.No. 4.8e-292; Indels 0; Gaps 0; Matches 1736; Conservative
Db	41130	CATCCTCCTAAGAAGAAAATCCAAATAGTTTAAGGAAATTTTGTGTAATTCATAGTCT 41071
Qy	1381	TTTTCGPTAACACAGTGTCTAATGATGATATCGTCAATCATATTCCTCTTTGGAACAACAAA 1440
Db	41070	TTTTCGTAAACACAGTGTCTAATGATGATATCGTCAATCATATTCCTCTTTGGAACAACAAA 41011
Qy	1441	AAGATCGTTTTTGAAAAATTAGTAGGGCACTAAAGTCGCATTTGTGTCCGTGCGAAA 1500
Db	41010	AAGATCGTTTTTGAAAAATTAGTAGGGCACTAAAGTCGCATTTGTGTCCGTGCGAAA 40951
Qy	1501	TCTAGCGTCTGTGCATCCACAAATAAGTGTGTGATTCGAGCTTCCAAGATTATAATCTT 1560
Db	40950	TCTAGCGTCTGTGCATCCACAAATAAGTGTGTGATTCGAGCTTCCAAGATTATAATCTT 40891
Qy	1561	TTTTAGATGGGTCAATGAAGATTTCTAATCTCGTATAGAGTGATCAATATATTTCTAA 1620
Db	40890	TTTTAGATGGGTCAATGAAGATTTCTAATCTCGTATAGAGTGATCAATATATTTCTAA 40831
Qy	1621	CATATAGTCTGT 1680
Db	40830	CATATAGTCTGT 40771
Qy	1681	TTGTGACACCTAGACAAATCATATATAGTGTACGTGAGTGTGACACCAAATATGGTC 1740
Db	40770	TTGTGACACCTAGACAAATCATATATAGTGTACGTGAGTGTGACACCAAATATGGTC 40711
Qy	1741	CAAAATATAATTTAAATTTGGCCACAAACAAATTTTACAACAAAATTCACAAACATGC 1800
Db	40710	CAAAATATAATTTAAATTTGGCCACAAACAAATTTTACAACAAAATTCACAAACATGC 40651
Qy	1801	ATCGTTTCAAAATTTTATTTATTCATAGCGCTTATTTGTTCATTTGTAATATTCGTGTTAA 1860
Db	40650	ATCGTTTCAAAATTTTATTTATTCATAGCGCTTATTTGTTCATTTGTAATATTCGTGTTAA 40591
Qy	1861	CTCAGTGCAGAAATTTTAAATTTTCAAAGAAAGAACATTTTGTGATATAAATAACATTTTA 1920
Db	40590	CTCAGTGCAGAAATTTTAAATTTTCAAAGAAAGAACATTTTGTGATATAAATAACATTTTA 40531
Qy	1921	TGGAACCCACCGGTAAAGCTCGATGATTTTGAGTTTGTAGTTTGTGTTTGTGAAATCAT 1980
Db	40530	TGGAACCCACCGGTAAAGCTCGATGATTTTGAGTTTGTAGTTTGTGTTTGTGAAATCAT 40471
Qy	1981	TAAAGACCTACATTTGATGCCCTCATTTTAAATTAATTAAGTAATCAACATGATGATTA 2040
Db	40470	TAAAGACCTACATTTGATGCCCTCATTTTAAATTAATTAAGTAATCAACATGATGATTA 40411
Qy	2041	GTTCCACCAAGACGCTCTTATGGCTATTAAAGATCAGACGAAGATGACGGGGTCAAT 2100
Db	40410	GTTCCACCAAGACGCTCTTATGGCTATTAAAGATCAGACGAAGATGACGGGGTCAAT 40351
Qy	2101	TAAGAGCTCTTATTAATTAACCAATTAATCAACTTAATGCTTAATTAATCAAG 2149
Db	40350	TAAGAGCTCTTATTAATTAACCAATTAATCAACTTAATGCTTAATTAATCAAG 40302
	RESULT 7	
	AF242313	LOCUS 1741 bp DNA linear PLN 06-AUG-2000
	DEFINITION	Arabidopsis thaliana unknown sequence.
	ACCESSION	AF242313
	VERSION	AF242313.1 GI:9719364
	KEYWORDS	' Arabidopsis thaliana (thale cress)
	SOURCE	Arabidopsis thaliana
	ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
	REFERENCE	1 (bases 1 to 1741)
	AUTHORS	Mollier P., Hoffmann B., Pelletier G. and Chwetsoff S.
	TITLE	Direct Submission
	JOURNAL	Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr, Versailles 78026 France

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Tracheophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1741)
Mollier, P., Hoffmann, B., Pelletier, G. and Chwetzoff, S.
Direct Submission
Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
Versailles 78026 France

Db 841 TATAGCAAGAAATATATGATGTAGATGTTGTCTCTTTGTAGTTACTCTTTTGTTCGCTA 900
Qy 1309 GTCCCTTTCTCTCATCTCCCTATGAGAAATATCCAAATAGTTTAAAGAAATTTTGTGT 1368
Db 901 GTCCCTTTCTCTCATCTCCCTATGAGAAATATCCAAATAGTTTAAAGAAATTTTGTGT 960
Qy 1369 AATTCAATGTCCTTTTCTGTAACCAAGTTCATAGTAGCTATCGTCAATATCTCTCTTT 1428
Db 961 AATTCAATGTCCTTTTCTGTAACCAAGTTCATAGTAGCTATCGTCAATATCTCTCTTT 1020
Qy 1429 GCACCAACAAAGAAATGTTTCTGTAACCAAGTTCATAGTAGCTATCGTCAATATCTCTCTTT 1488
Db 1021 GCCCAACAAAGAAATGTTTCTGTAACCAAGTTCATAGTAGCTATCGTCAATATCTCTCTTT 1080
Qy 1489 GTCTGTGCGAAATCTAGCGTTCTGTCATCCCAAAATAGTTTCTGAGCTTCGAA 1548
Db 1081 GTCTGTGCGAAATCTAGCGTTCTGTCATCCCAAAATAGTTTCTGAGCTTCGAA 1140
Qy 1549 GATTATAATCTTTTGTAGGCTCATGAAGATTTCTAACTTCGATACGAGTGTATCCA 1608
Db 1141 GATTATAATCTTTTGTAGGCTCATGAAGATTTCTAACTTCGATACGAGTGTATCCA 1200
Qy 1609 TATAATTTCTAACATATAGCTTCTTTTGTAGGCTCTGCTTTTGTAGACCAACCC 1668
Db 1201 TATAATTTCTAACATATAGCTTCTTTTGTAGGCTCTGCTTTTGTAGACCAACCC 1260
Qy 1669 CTTCGTAATGTTTGTGCACTTAGACAAATCCCAATACGTTACGTCAGTCGAAGTTGC 1728
Db 1261 CTTCGTAATGTTTGTGCACTTAGACAAATCCCAATACGTTACGTCAGTCGAAGTTGC 1320
Qy 1729 ACCAAATGTCGCAATATATATTTAAATTTGGCCACAAAAACAATTTTCAACAAAT 1788
Db 1321 ACCAAATGTCGCAATATATATTTAAATTTGGCCACAAAAACAATTTTCAACAAAT 1380
Qy 1789 CAACAAACATGATCGTTTCAAAATTTATTTTCAATGCGGTTATTTGTTTCATGTAAA 1848
Db 1381 CAACAAACATGATCGTTTCAAAATTTATTTTCAATGCGGTTATTTGTTTCATGTAAA 1440
Qy 1849 TATTCGTGTTAACTCACTGACGAATTTTAAATTTTCAAAAGAAACAATTTTGTATATA 1908
Db 1441 TATTCGTGTTAACTCACTGACGAATTTTAAATTTTCAAAAGAAACAATTTTGTATATA 1500
Qy 1909 AATTAACATTTTATGGAACCAACCGTTTAACTGTCGATGATTTTGTAGTTTGTGCTT 1968
Db 1501 AATTAACATTTTATGGAACCAACCGTTTAACTGTCGATGATTTTGTAGTTTGTGCTT 1560
Qy 1969 TTGTGAATCATTAACGACCTTACATTTTATGTCCTTACTTTTAAATTAATTAAGGAATCAA 2028
Db 1561 TTGTGAATCATTAACGACCTTACATTTTATGTCCTTACTTTTAAATTAATTAAGGAATCAA 1620
Qy 2029 CATGATGATTAAGTTCACCAAGACGTCCTTTATGCTTATTAAGAGTCAGACGCAAGAT 2088
Db 1621 CATGATGATTAAGTTCACCAAGACGTCCTTTATGCTTATTAAGAGTCAGACGCAAGAT 1680
Qy 2089 GACCGGGCTTATTAAGAGCTTATATCAACCAATTAATTAATTAAGTTCATTAATCA 2148
Db 1681 GACCGGGCTTATTAAGAGCTTATATCAACCAATTAATTAATTAAGTTCATTAATCA 1740
Qy 2149 G 2149
Db 1741 G 1741

RESULT 8
ATH553603/c 317 bp DNA linear PLN 29-MAR-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, left border, clone
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone
362F10.
ACCESSION AJ553603
VERSION AJ553603.1 GI:29369764
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
REFERENCE
AUTHORS Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G.,
Lepiniec, L., Caboche, M. and Lecharny, A.
T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)
2
JOURNAL
MEDLINE 22363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 317)
AUTHORS Balzergue, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.infobiogen.fr).
FEATURES
source
1..317
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultiivar="Wassillewskija"
/db_xref="taxon:3702"
/clone="362F10"
misc_feature 1..317
/note="T-DNA flanking sequence
left border"
ORIGIN
Query Match 11.7%; Score 251.8; DB 8; Length 317;
Best Local Similarity 99.2%; Pred. No. 4.6e-34;
Matches 253; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1753 AAATTTGGCCACAAAAACAATTTTACAAACAATTCACAAATCGATCGTTCAAAT 1812
Db 317 AAATTTGGCCACAAAAACAATTTTACAAACAATTCACAAATCGATCGTTCAAAT 258
Qy 1813 TTATTTTATTTCAATGGCGTTATTTGTTTCATTTGTAATAATTTCTGTTTAACTCACTGACGAA 1872
Db 257 TTATTTTATTTCAATGGCGTTATTTGTTTCATTTGTAATAATTTCTGTTTAACTCACTGACGAA 198
Qy 1873 TTTTATTTATTTTCAAAAGAAACAATTTTGTATATAATAAATTTTATGGAACACCGG 1932
Db 197 TTTTATTTATTTTCAAAAGAAACAATTTTGTATATAATAAATTTTATGGAACACCGG 138
Qy 1933 TTAAGCTCGATGATTTTGAGTTTGTAGTTTGTGTTTGTGAAATCATTAACGACCTACA 1992
Db 137 TTAAGCTCGATGATTTTGAGTTTGTAGTTTGTGTTTGTGAAATCATTAACGACCTACA 78
Qy 1993 TTTGATCCCTCATTA 2007
Db 77 TTTGATCCCTCATTA 63
RESULT 9
AL773135
LOCUS Arabidopsis thaliana transposon insertion STS SM_3.1249, sequence
DEFINITION tagged site.
ACCESSION AL773135

```

VERSION      AL773135.1  GI:21535322
KEYWORDS     STS; STS, sequence tagged site.
SOURCE       Arabidopsis thaliana (thale cress)
ORGANISM     Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE    1
AUTHORS      Clarke J.H., Bowles B., Carter J., Hart D., McCullagh B.,
MURPHY G., Langham S., LeGrys C., Jones J.D.G. and Bevan M.
Unpublished
JOURNAL      Direct Submission
REFERENCE    2 (bases 1 to 295)
AUTHORS      Clarke J.H.
TITLE        Submitted (18-JUN-2002) Clarke J.H., John Innes Centre, Colney
JOURNAL      Lane, Norwich, NR4 7UU, UK
COMMENT      AT denotes an activation tag dissociation transposon within a
single line. ET an enhancer trap dissociation transposon, GT a gene
trap dissociation transposon, MT a mis-expression enhancer trap
dissociation transposon, SM a defective suppressor mutator
transposon. _3 denotes a sequence derived from the 3' end of the
transposon, _5 denotes a sequence derived from the 5' end of the
transposon. BESC GARNET, ATIS project
On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
code: N55909.
FEATURES     Location/Qualifiers
source       1..295
             /organism="Arabidopsis thaliana"
             /mol_type="genomic DNA"
             /variety="Columbia-0 NASC stock code N1092"
             /db_xref="taxon:3702"
             /clone="AC012193"
             /notes="Derived from superpool 12.77 NASC code Unknown"
STS          1..295
             /standard_name="SM_3.1249"
ORIGIN
Query Match      6.1%; Score 132; DB 11; Length 295;
Best Local Similarity 76.4%; Pred. No. 3.3e-13;
Matches 162; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 321 ACATCAGTATCTCAAGTCGAATAACAATATCCATAGAGTGGTATATTGTGAAAAAAA 380
Db 1 ACATCAGTATCTCAAGTCGAATAACAATATCCATAGAGTGGTATATTGTGAAAAAAA 60
Qy 381 AAAAAAAGGTGTATCTGGTATATACATATACACCGTCTCGAATTGCCTCAACAATT 440
Db 61 AAAAAAAGGGGAATCCGGGTTTAAACCAACCCCGGCCCAATTTCCCCCAACATTTT 120
Qy 441 CTAGGAGAAAATGGACGTCTCTTTGGTTTATTTTATCTTAATACATCTCTATAT 500
Db 121 CTGGGAAAAAGGGGCCCGGCCCTTGGGTTTTTTTTTTCTTAAAAACCCACCCCTATT 180
Qy 501 TTTAAACACATTCGATGCTCGCTTAAATTTTCG 532
Db 181 TTAACCCCTCCAGGCCCTTAAATTTTG 212
RESULT 10
AC007289/c 43982 bp DNA linear PLN 27-FEB-2002
LOCUS      Arabidopsis thaliana chromosome 2 clone F16J10 map RNS1, complete
DEFINITION Arabidopsis thaliana (thale cress)
ACCESSION  AC007289
VERSION     AC007289.9  GI:20198243
KEYWORDS   HTG.
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana (thale cress)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE    1 (bases 1 to 43982)
AUTHORS      Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Shen, M., VanÅken, S.E.,
Barnstead, M.E., Mason, T.M., Bowman, C.L., Bowring, C.M.,
Benito, M.-I., Carrera, A.J., Creasy, T.H., Buell, C.R., Town, C.D.,
Nierman, W.C., Fraser, C.M. and Venter, J.C.
Unpublished
JOURNAL      Direct Submission
REFERENCE    2 (bases 1 to 43982)
AUTHORS      Lin, X.
TITLE        Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL      Medical Center Dr., Rockville, MD 20850, USA
REFERENCE    3 (bases 1 to 43982)
AUTHORS      Town, C.D. and Kaul, S.
TITLE        Direct Submission
JOURNAL      Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA. cd-town@tigr.org
COMMENT      On Apr 18, 2002 this sequence version replaced gi:6598749.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES     Location/Qualifiers
source       1..43982
             /organism="Arabidopsis thaliana"
             /mol_type="genomic DNA"
             /cultivar="Columbia"
             /db_xref="taxon:3702"
             /chromosome="2"
             /map="RNS1"
             /clone="F16J10"
misc_feature 1..5724
             /notes="overlap with BAC clone T20G20
             (AC006220:104629..110352)."
misc_feature 1
             /notes="17532 nt before this point were not included in
             the submitted sequence, due to overlap with another BAC"
repeat_region complement(1434..1472)
             /rpt_family="AT_rich"
repeat_region complement(1907..1945)
             /rpt_family="AT_rich"
gene         complement(2734..3407)
             /gene="At2g05435"
             /note="synonym: F16J10.1; pseudogene"
             /pseudo
repeat_region complement(3455..3489)
             /rpt_family="(TA)n"
repeat_region complement(3808..3836)
             /rpt_family="AT_rich"
repeat_region complement(3859..3914)
             /rpt_family="Rf:ATREP9|ATREP9 ATREP9 repeat - a
             consensus."
repeat_region complement(3978..4108)
             /rpt_family="AT_rich"
repeat_region complement(4754..4787)
             /rpt_family="AT_rich"
gene         complement(5500..6147)
             /gene="At2g05430"
             /note="synonym: F16J10.2; predicted by genefinder"
             /complement(join(<5500..5949,6031..>6147))
             /gene="At2g05430"
             /complement(join(5500..5949,6031..6147))
             /gene="At2g05430"
             /codon_start=1
             /product="hypothetical protein"
             /protein_id="AAW15480.1"
             /db_xref="GI:20198245"
             /translation="MSKISRSKSTLSHHGTAYDELDNDEIVNIKGFWRFSKJGLAK
             RVPEKHPETTKFCLKSEFAKTVLTALLIDIMNLPQOOLPEALKEALNTLIDL
             EAAGFKLDWLKRKLEIRVTKKAKNTTTRRELDRIKIQHLELSVLQEBMKKEQE
             AMCDKPEYLLDSICLNEVTCDSSEPTK"
             /complement(7260..8700)
             /gene="At2g05420"
             /note="synonym: F16J10.3; predicted by genefinder"
             /complement(join(<7260..7640,7857..8077,8329..8510,
             8591..>8700))
             /gene="At2g05420"

```

```

CDS
complement(join(7260..7640,7857..8077,8329..8510,
8591..8700))
/gene="At2g05420"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32895.1"
/db_xref="GI:4914358"
/translation="MKQVDSKTIITWVIFNSLSQASIHSDQFVVGDCRWLKAYPK
GNEKATYIAYRANLALXNVANSKSPFPGWTRHTKFTSLVNOSEKLSKLTQSHW
FDHKSITSGFFAMIPLEIHLNTHNEGLFVAKVEVLGVKLDVSKSSPVMET
IDVNGFQVLPQESVNRILPAKHILDIASKRPKPKYMTAYMVLISLTKTLCOSQDIL
SNDNISGAALUTYLRAGFKLDLWLEKKHGEIKKKKEEASLKLQMEKQIFNEAQ
IYKVKLAARPLSLNEDNV"
repeat_region
9262..9302
/rpt_family="AT_rich"
complement(9910..9970)
repeat_region
/rpt_family="AT_rich"
complement(10006..10019)
repeat_region
/rpt_family="TAAA)n"
10066..10101
/rpt_family="AT_rich"
complement(14048..15111)
gene
/gene="At2g05410"
/note="synonym: F16J10.4; predicted by genscan"
complement(join(<14048..14464,14624..14847,14955..>15111))
mRNA
/gene="At2g05410"
complement(join(14048..14464,14624..14847,14955..15111))
CDS
/gene="At2g05410"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32896.1"
/db_xref="GI:4914359"
/translation="MAYPKGINKAHDSSFLNVPDNLSPFTGMRRAKVFSLVNOG
SEKLSQKVTQHWFKQAFQVFPVMTHTELNAKMGFLVNGELKVAKIEVLGVGK
LDVSKSSPIMKIDVNGFQVLPQESVNRILPAKHILDIASKRPKPKYMTAYMVLISLTKTLCOSQDIL
SNDNISGAALUTYLRAGFKLDLWLEKKHGEIKKKKEEASLKLQMEKQIFNEAQ
IYKVKLAARPLSLNEDNV"
repeat_region
complement(15873..15972)
/rpt_family="AT_rich"
complement(16262..17096)
gene
/gene="At2g05400"
/note="synonym: F16J10.5; predicted by genscan"
complement(join(<16262..16672,16854..>17096))
mRNA
/gene="At2g05400"
complement(join(16262..16672,16854..17096))
CDS
/gene="At2g05400"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32897.1"
/db_xref="GI:4914360"
/translation="MFCDSAEQWFDKNAPDWGFTMI SLELHAEGLRNGELTV
VAKVEVLGVKLDVSKSSPIMKIDVNGFQVLPQESVNRILPAKHILDIASKRPK
NPYLKAYMVLISLTKTLCOSQDILSNDNISGAALUTYLRAGFKLDLWLEKKHGEIKKKKEEASLKLQMEKQIFNEAQ
IYKVKLAARPLSLNEDNV"
complement(17875..21892)
gene
/gene="At2g05390"
/note="synonym: F16J10.6"
complement(join(<17875..19791,19831..20615,20671..>21892))
mRNA
/gene="At2g05390"
complement(join(17875..19791,19831..20615,20671..21892))
CDS
/gene="At2g05390"
/codon_start=1
/product="putative retroelement pol polyprotein"
/protein_id="AAD32898.1"
/db_xref="GI:4914361"
/translation="MEATLRVHKVWETIDPQSDMDKMDMARALLFQSVPESTILQVG
KHKTSKAMWAIKTRNLGAEKVEAKLQTLMAEFDRLNMDNETIDFVGRKISEITK
SESLGIESIEKVLKSLSPKKYTHIIAALQILDNLNTGTFIDVGRMKYEDRV
CDEDSPFQGLKMYANSESYTRGGRGSSGRGGYQOORQKQVYKVCRC
DKTGHVASECLDRLLKILKAQEQANNEDDEIESLMHVEVYLNERSVKPPEFACS
DWSYLDNGASNLGNLQFSLKNEWITGKVRGDDSDIDIKGKSIVLITKGGRK
TLTDVYFIPDLKSNILSIGATGACDVRMKDQDLTHDREGCLLRATRSRNLVYK
DLNVNPKLQLEAATNVRKELVIGISNIPKEKETCGSLGKQARQPPFKATTYRAS

```

```

OVLVLVHGDLCGPITQSTTAKKRYILVLIDHTRYMWSMLLKEKSEAFKFRDFKTKV
EOESGVKIKTPTDKGFVSOEFODFCAKEGINRHLTAPTYTPOONGVVERNLTLLG
MTRSLHKMKMENVLGEAVRHSYILNRVGTGRSLNQTPYFVFKORKENVEHLRVFG
CIGYAKIIGPPLRLKLDLRSKMLYILGTEPESKAYRLDPTNRKILKNWNSDSETRDIS
GTFSLTGEFGNGIOESDDIETEKNEBESSEENHEENBOEIQIDAEETQPSHA
PLTPTRRSTQVGPKNYLDLDDYVLAIEGEQVLAINDPEWDKFEANKLKEWRDACK
BEILIERKNTWSLIDLVRVFKIGLKWVFKIKRNSDGSINKVKARLAKGVYVRHGI
DYDVFARHATIRVILAIASAANGWEVHLVDKTAFLHGLREDVYVYTOEGFTNK
DNEGKVKYLKALYGLKQAPRAWNTKLKLOELNFVKSKSPSVRRRBEKKLLIVA
TYVDLLVTGSSDLILCFKDMAGFEMSDGLQTYTYLIGIEVLHKNNGIILRERYA
MKIIEAGMSNCNPLVILPMAAGLELCKAEEKCIITERYRMIGCLRIVHTRPDLISY
CVGLSRYLQFPRESHGALQVRLYKLTGSHGLYLRGFKSLGVGSDSHSADLD
DGKSTAGHIFYLHQCPITWCQOQWALSCAEAFMAATEAKAKAIWLQDLFAEVCG
TTSEKVMIRVNDKNSAIALTKNLVPHGRSKHRRVYHRECVENNLVVDHVPVGEQR
ADILTKPLGRIKFREMRELVGQVLEDDFKLKGFEVLGSMKSN"
complement(22479..22502)
/rpt_family="AT_rich"
complement(24570..25781)
/gene="At2g05380"
/note="synonym: F16J10.7; supported by cDNA:
gi_15146251.gb AY049267.1"
complement(join(<24570..25058,25678..>25781))
/gene="At2g05380"
complement(join(24781..25058,25678..25750))
/gene="At2g05380"
/codon_start=1
/product="expressed protein"
/protein_id="AAD32899.2"
/db_xref="GI:20198244"
/translation="MAKSTLLLLGLFAFLFIVSEMAAAGTVKSESEETVKPEHGGGF
GDNGGGRYQGGGGHGGGGYQGGGRYQGGGRYQGGGGYCRHGGCYKGVHGCRC
CSYAGEAVQTQSGH"
repeat_region
complement(24878..24990)
/rpt_family="GGA)n"
repeat_region
complement(25118..25153)
/rpt_family="AT_rich"
repeat_region
complement(26299..26329)
/rpt_family="(A)n"
repeat_region
complement(28917..28939)
/rpt_family="AT_rich"
repeat_region
28929..29834
Query Match 4.4%; Score 93.8; DB 8; Length 43982;
Best Local Similarity 89.4%; Pred. No. 7.8e-07;
Matches 101; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 74 TTCTACCTTTTGGCATGGACCATGTTTATAGGATTTACTTGTGAATTCCTGAATCT 133
Db 34629 TTTTAACTTTTGGCATGGACCATGTTTATAGGATTTACTTGTGAATTCCTGAACCT 34570
Qy 134 TTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACAGTTCT 186
Db 34569 TTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACAGTTCT 34517
RESULT 11
PFA929353 343050 bp DNA linear INV 29-JAN-2003
LOCUS Plasmodium falciparum strain 3D7, chromosome 5, segment 3/4.
DEFINITION AL929353 AL844504
ACCESSION AL929353.1 GI:23504644
VERSION AL929353.1
KEYWORDS
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D.,
Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K.,
Buckee, C. O., Burrows, C., Cherevach, I., Chillingworth, C.,
Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C.,
Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J.,
Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N.,
Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holtroyd, S.,

```

Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Lark, N., Lawson, D., Lennard, N., Line, A., Maddison, M., Mclean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A., Rabinowitz, E., Rajandream, M.A., Rutter, S., Rutherford, K.M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and Barrell, B.G.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 343050)
Devlin, K., Baker, S., Davies, P., Mungal, K., Berriman, M., Pain, A., Hall, N., Bowman, S., Churcher, C., Quail, M. and Barrell, B.

TITLE
JOURNAL

Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/p_falciparum.

COMMENT

FEATURES
source

```

1. 343050
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="5"
complement(148..1134)
/gene="PF0795c"
complement(148..1134)
/gene="PF0795c"
/notes="Similar to Mus musculus nif-like protein nifL
TR:CA69078 (EMBL:AJ344340) (348 aa) fasta scores: E():
7.1e-16, 36.310% id in 168 aa
Similar to Dictyostelium discoideum development protein
dg1148 dg1148 TR:Q9XYLO (EMBL:AF111941) (306 aa) fasta
scores: E(): 2.6e-16, 37.267% id in 161 aa"
/codon_start=1
/product="nif-like protein, putative"
/protein_id="CAD51523.1"
/db_xref="GI:23504645"
/db_xref="SPTREMBL:Q8I3U9"
/translation="MKKRFPERCLQILKIVFVMTIWKIKLIVKVLGSLFLRDSF
HFYTKRRICPTMTVLDELTYICTKKRHYQKEDVDVLNGKLYPLVCKRPYI
DLFSSLYPVEIITFTAIKSVADTVLNTIDVHYIDKRFYEDCYEMNEKLYIKNL
TNLKEISLIIDDSNISGFOYDPNFPPIKMGQDLNDELHLLHLPFLNLRKLDI
RSLCSFLIKQKOVSKLLCTSSSKHILKYLTDNKSGLYRSLAFQALNYLKNKFR
YASKKQWNLGRKNNKLSYFSLSKLSSDDEHKKHKKRKLNNRKLRL"
join(3549..6704,6817..6915)
/gene="PFE0800w"
join(3549..6704,6817..6915)
/gene="PFE0800w"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51524.1"
/db_xref="GI:23504646"
/db_xref="SPTREMBL:Q8I3U8"
/translation="WRKYSIKFINSRLYSRRYSRNOFYNSFKCNKNEQRIKSKND
DPLDESNNVEFDEYISPSFDINTDDKKWKELEYNKEIKTKDQINPTTFP
NQCKTKHNTSGNSYEKLQIHNKHIDNNISNVTSNNTPHNVETVEKENIYE
DDINIPLKNSDDITLQGGIILKKLIYLSYKSLHTYNNKKGILGAYLNSILLBRGST
DKDGVKTRKEENKEQNEEDKEYADILYDKNKERTETENRGGHHIYQLQ
HGFNNIDPESLANKLQKLVNTFTAGYTYNTTFKLIIVENTQSLTYNNLFTLNL
LHLNMKNMRELDSEVILMKRKKSDQEDILQANLNKKNKKINDIENNIKLVCH
YINLNFSLYFCEDYDLJEYINNKNVDLFCFTLQCFHSLTKHVLDKIYFLCLK
KIHFNFEELKMLKICFRFCCKYSKYFVLSIDIIKKYIHLNMLNFVQLLKITN
AFNKNKYVELKGLAEKIMRVGNIQREKKKKQGYLNNTNNTSVNVDMMYNNRHK
KKSSSRNCPDYDTLGLGANSVCFVNFENSHVNTIYEDNIIKIQLCKLYEYNNRKS
BHVTCKNVIYEMINENMECKLIEDIYVSYVCFSTYNNRKLILYNNFLDIYKSN
ELMNAKNCILWIYPLSLCKISWHRNVIYELFGYIKDNYILNRLSVFQVLKLLSSIV

```

gene

CDS

```

KNYIDEEVYKILIEKLYKEWDI IKKLLIDISTFLMCCAYINIIYKPLFDSAYNLLIID
LINKNVPILNNVMYKCNFVNITWGFIVANVHKHTNEDQILNITLTKRNPNENEAFK
LHQJSDSCFKSEIPKTLINCLDITLTYAYCTHEKCKLLRNNTIYKEDVAKVWKKI
HDELINLNSFSHISDFQVEPYHNSFYIIDICLNQTSKIGWISFOKEVIMRLTKSS
WDFMTGIVTLQWRIILHAHGWKIIIPINAGEWIQLNFDQKKNLLSEYFKQYSIQI"
join(7484..14348,14522..14859)
/gene="PFE0805w"
join(7484..14348,14522..14859)
/gene="PFE0805w"
/BC_number="3.6.3.-"
/notes="Similar to Plasmodium falciparum probable
cation-transferring ATPase 1 SW:ATX1 PLAFA (Q04956) (1956
aa) fasta scores: E(): 0.97.919% id in 1970 aa
Large N- and C-terminal extensions
aa 50-72, 286-308, 334-356, 361-380, 708-730, 2173-2195,
2207-2229, 2263-2282, 2302-2324 and 2334-2356"
/codon_start=1
/product="cation-transferring ATPase 1"
/protein_id="CAD51525.1"
/db_xref="GI:23504647"
/db_xref="GOA:Q8I3U7"
/translation="MYKNIFFPHDNIKETYKYVYVYHTNLLIHDEYRISVIECYKKK
KIVTILHVFCILSLGLPLILSWPPIFVRLHSACTINECEKVLITPKDGNKYIK
KIKIKFNHVCYVNDVDFSEDQSMVNDKEONKNPNVFSRDVFRDQNMNER
KUNINLKHILKLSHINLKNILKSGHINLKNILKSGHINLKNILKSGHINLKNIL
NFKDSIYPCDEILHVDNINLNNLINSQFPFETTEYNKSRLOQNNRIYIIFELDI
RYVYIKKNI FVELYFNIFPFSYMHLMCTGLRNEKLINDRKLILYGCNLLINLKSDFI
ILIFKEMPPFFIQIFAMIVSLDNIYIETISILFITSIIILEKNTIKNKKIKN
MLNVTCPINVRVNTYIISSELPVDEIYKNNMTIPDITIIILSGVTSBEMLTG
ESVPIKEOLPREGNALINKNKVESDEKDYLRVYNNHASINMIKRNHLISETLKG
KREYKSNTHDLCSNKLVCYNNYDDEVKNNKMDYNNNNKKKKNILNLFVKGT
YINNSOLDYDKIGVNI FDDVNNMKHKNQNRNINYNKDTNNLNNKHRYIVDCIL
KKEVALSQNNKIIISNEDINKMYLIGTYVLSLYNINKIKYNNKEENRILGLVKTGF
ITTKGIVNNILYHKKELNINDSYKFLILIIYAFSFFILITLYSNNEYNNHII
IKCLDIIITDAIPALPTLTVGISIAISRLKKFSISCLCPHKINTAGQINTWVFKT
GTUENLQFIHITONKKNKNSDFIHIKEMNTESYIHSKODNMTHNKSIISEYI
IKDNKMLHTSSKKSITKERSNPLAQITKSCLLKHVYIKEKKKYVNTYNTCNDLHI
NDTCSYLLNSETDAYCEYNIIDHLCIDNKKMDINSKNELMGKYSKNELMGKYSK
NELMGKTIKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSKNELMGKYSK
KYSKNELMGKTIKNGQVDNTIYHNMCDNDYDPCDNCNNCDNTIYHLEVHNIN
KDSFNIPPEKNSYNNISEHIKINYPLPEALACCHTLKSVNNKMGDVLIELMFPF
TNCMDLINNSFTIIEKKKNCSDYDFQIDGDKNIGANDERCHLNINLVSYLKRFEP
OSRLQMSVLKSTYGNNDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDNDDN
NIFOKSPEKIZELCISKSTPNYDEILNKYTKGWEILSISYKRVKSKNILLNVKR
SFVESNIIIGDVI PVVGNKNNKQKGLVYNNKNDTYLKGDKTICIDNEFTSIQSQMS
NSNVEISIIIGDVI PVVGNKNNKQKGLVYNNKNDTYLKGDKTICIDNEFTSIQSQMS
DNICGDNICGDNICGDNICGDNICGDNICGDNICGDNICGDNICGDNICGDNICG
NGNINTYDNIYGDNYLDYCPTEYHKTYYNNSITLYRNFLYKKNKKNYKNIISIL
YEHRTNDIQDKLCLDILINKDPRNVILVTKAFIFLKKKFPFLPYEYECNKIVHY
IMKHKHKKIKNIIINNSNLYHYNIIDTVKRMKYMCFNKLKLYIOQLKLYNLH
NLYKXKKNYIYDIDVHLI GNNNNNNKSKKELPLANKKHIFKNSNDNITEN
TYTSNNIYLSKYKYVHHKNNYYPDSCTNLRRKNSLFINLKKIYIYKXKLYLOCLLK
HDNYKVELPRIKDINSYQMESIKTRNFTHSISEQFAISNLISLYIINKDNNVYI
KNKYVKNKYVKNKSNICKNKYI CNKYIYKNKYIYKNKYIYKNKYIYKNKYIYKN
SKKELAFSNIIRRHKLKKNKKNKRYKNHVNNTSKGHIILNNCTGFKKKDYS
LKNKYRIVNKGWMLKNDVYDHYHVNLTDMYRGTQYCGSKKNKNTIYMMNNNLIK
KINRFLHLIVDKCKNI CHKYTDIKNISIEYIYIYIYIYIYIYIYIYIYIYIYIYI
KPLNNYSVCMGCGANDCLACADIGISLCNNNSISGSSFTSNKCLHSIYHILI
EGRASVNSQLFKFISLYSIMQCSQILYISNKLTDNQYIFIDIVITLPLISIFMC
WTSASEKLSKNIPIGLFSPPIISYQIIIIQLFVYIISLVLLNLSFYKYDKNNM
KEKSDDYLYKAQNTLYILCSFONLFWCISLNIKNKWRKSVNTIYAFIYMSFKLLM
NTCTIPFSEIFVGIWITDILLKQVLSITRPPFIRFLFTLFFNFLCSYSYEKVIK
YFKEBQKXKNNHNSIIPCSPOELDTTVFVTCN"
complement(join(15807..15912,16021..16198,16441..16612))
/gene="PFE0810c"
complement(join(15807..15912,16021..16198,16441..16612))
/gene="PFE0810c"
/notes="Similar to Arabidopsis thaliana putative 40S
ribosomal protein s14 f24k9.19 TR:Q9CAX6 (EMBL:AC008153)
(150 aa) fasta scores: E(): 2.9e-44, 81.457% id in 151 aa"
/codon_start=1

```

gene

CDS

Qy	818	TTACAAGGGTCAAAACCATATCCAAATGGTTCTCGCTATATATATATATACTAAAGATCCCC	877
Db	162378	TAAAAATATTAAATATTATATATATATTTTAAATATTATATTTATATATATATATTTAA	162319
Qy	878	TCCTGTGCTTGCTTTATTCGTCGATATATATCTAACTTAAATTTAGTTCCTAAAAATATAT	937
Db	162318	ATAATTTCTATTTATATATATATATATAAATAATTTTCTATTTATATATATTTTAAATAT	162259
Qy	938	ATGCTCCACCATGTT-TCTACTGACCTCAGTCCTAGTTAGCTATATATGGACATATGGA	996
Db	162258	TTTCTATTTATATATATATATAAATAATTTTCTATTTATATATATATAAATATTTTC	162199
Qy	997	AAATGAGCCCCAAAATTTTGAGAGGTTCTCTTCTCGCAACTTACTCTTATCTTACTCAT	1056
Db	162198	TATTTATATATATATATAAATAATTTTCTATTTATATATATATATAAATATTTTCTATT	162139
Qy	1057	GAGCTATGTTAAATATTTGCAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACCTAGA	1116
Db	162138	TATATATATATATATAAATAATTTTCTATTTATATATATATATAAATAATTTTCTATT	162079
Qy	1117	TAAAAAAACATGATAGACATTTAGTGTAAAACTTGAAATGTTTATTGAACTCTTTGGAA	1176
Db	162078	TATATATAATATATAAATAATTTTGTATTATATATATATATATATTTTATATTTATATAT	162019
Qy	1177	ACGTGGATGTTGTATGGATTAATTTTGAAGATATTTATATATTTGAGAGATCTTTTATATA	1236
Db	162018	ATATATATTTTATATAAATAATTTTATATATATATATTTTATATATAAATAATTTTAA	161959


```

/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(178. .211)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region 307. .327
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(A)n"
/repeat_region complement(328. .348)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region complement(378. .423)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(479. .503)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(513. .539)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(A)n"
/repeat_region 577. .633
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(A)n"
/repeat_region complement(714. .747)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region 841. .867
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(A)n"
/repeat_region complement(888. .918)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(1006. .1049)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(1075. .1120)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(1140. .1283)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region 1294. .1346
/repeat_region /rpt_type=tandem
/repeat_region 1298. .1346
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TATAA)n"
/repeat_region complement(1423. .1531)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(<1570. .2424)
/repeat_region /locus_tag="PF0490c"
/repeat_region complement(<1570. .2424)
/repeat_region /locus_tag="PF0490c"
/repeat_region complement(1570. .2424)
/repeat_region /locus_tag="PF0490c"
/repeat_region /codon_start=1
/repeat_region /product="hypothetical protein"
/repeat_region /protein_id="AAC71887.1"
/repeat_region /db_xref="GI:3845198"
/repeat_region /translation="MKEKNEIMDYLSCLDDVVDREKSKGKSLKSSSTKKSDYKK
SSIFSKRDHKKSGSGRSGFINKSGSFKKPYNNRLINKNNYKGRMFHNGR
DNFKGRGSGSRVDFNRKGSFKRFRISNRKSSVSKYRGNGSGMGRKSNFKAPTSR
TVVTKRNNYKTSAPKNNLNLSLYRKNRTFALNKRSPKPVGTIKSSVPRKRIKK
GLKKGSLKSTKSTSGSKFKPLNKYFLSKI KIVTSLNKIPSLKQKQNTENVLPESL
NNATTKKK"
/repeat_region complement(2150. .2188)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(2446. .2468)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region 2504. .2656
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TATAA)n"
/repeat_region complement(2697. .3167)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TAAA)n"
/repeat_region 3047. .3132
/repeat_region /rpt_type=tandem
/repeat_region complement(3047. .3132)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region complement(3258. .3420)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(3450. .3514)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region complement(3532. .3559)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(3689. .3720)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(4024. .4114)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region 4107. .4146
/repeat_region /rpt_type=tandem
/repeat_region complement(4115. .4146)
/repeat_region /rpt_type=tandem
/repeat_region /rpt_unit="(TA)n"
/repeat_region complement(4348. .4386)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(4403. .4450)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(4545. .4785)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(4937. .5001)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(5102. .5191)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(5246. .5275)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem
/repeat_region complement(5437. .5484)
/repeat_region /note="AT_rich"
/repeat_region /rpt_type=tandem

```

[illegible]

```
/gene="VAR"
/notes="Similar to Plasmodium falciparum variant-specific
surface protein var-2 TR:Q26033 (EMBL:L40609) (2664 aa)
fasta scores: E(): 2.1e-94, 38.066% id in 2782 aa"
/codon_start=1
/product="erythrocyte membrane protein 1 (PIEMP1)"
/protein_id="CAD51367.1"
/db_xref="GI:23504491"
/db_xref="GOA:Q81495"
/db_xref="SPTREMBL:Q81495"
/translation="MGPPGIGTQGETAKHMFDRIGKOVYEVKNEANYISEGLK
SQATLLGERSVSLCOLVEDYRSKANGDVKYPCANRSPVRFSDESRQTYNRKID
NETDNRAGGACAPRRRLHLDYNELEMGKTSITKHDLAEVCMAGVEGDSIKTHYI
KHNTNDSAAELCTELARSFADIGDIRGDLYLGDILKQKQKKTEREKLEENLKRI
FGKTHEDVNGKEVLTRYKDINDPEFKLREDWTANRATVVKAIITCHAGESADKVF
RNTCNSDSVTFSPQNKYRCNGDKPEDKANVDPTTYFDYVQYLRFWEAEDDFC
RIRKHKLNAKEQCRQKYSQTDYRCNRNGVDCTQIRGNILVSDSECTCSVUCVP
FVKIENKLLFEFGKQKYTEIEKANGSTYGTIRQYGTINMYKIRDFQOQLQSGY
GDVNAFLELLNKETICDHPKVBKSDIDFNEGTEKTFPSHTEYCETPCWATKKKGID
GNWEEQYEEGCENYLMKPIDESKSTIDILLVDKDTSGTVMVEKGLGCLNDSKRTVM
ENWQCHVEKKSQYEDGDKDYCVLKDDKKKPEHRTIKSYTTLFNNINEMLKDSIDW
SKEKLTINNEKPTNIRECKSCDKCFKMWVQKEQWKELEHYEKENFSGDFGPRI
SPYVTLFNLQYSLYEMLRKYAAQOEIQLIEKNKNFEVKEDDINSITKFLQOE
KGIATKCEKOECKQKQOROKQPADKVVSRSGASPDTPATPKATEEKEEBEBE
EEDLDGSESEPEQAVDQDQGETTEKKVPATBERGSPKBTITPEKSDVCKIVAE
LFSVDNLKCACTQKYGNNRGLWKCIPTSGGEKATGSGESTGSDATGGSICVPP
RRRLYVPLTTFWASGGTQVETQASGNTETSOVGETTPQGQTPSESEAQTASQDP
SEKLTAFIESAAVETFLMDRYKENTKQSVLPLELPINGDTISDNNPEKLLKVG
KIPDFLRLMFTYLGVDYRDLIFSSNSTSDTGTETPSSNDNLKNIIVIEAGTDEKDK
QKMEIOQKIDILKQSGTTPPTPVTHSPSSGTPPSSWMENNAKHIWKGVCALTYKD
NTNGGPTQNDLAKALWDDKTNPEKSEYKIDQVELKEENSGTGVTKGHAEPGNT
PLKSFVARPPYRLYEWGETFCRERTRLGKIKECKVDSDSGSGGVTKRQYSGDG
EDCKDHEDPTPLDGLSSCPKSSYRKMIKKOYEKQKAYEQKKNCEKESBG
VENKSDICQLQNLKNYGSIDFSLDLKKGCSKTYNESGDDNNKIFDETEKTFGH
QNYCGPSEFKINCKGKNDKQECNGKKTIDAQNFQMGOTAKSEFVWVSNDQTT
KFEHGLEDGNCAGIFGKIENKWKCDVGVYVCKPKNGKNGKECTYIIQIRALVR
WVEYLEDYNNKIRKLPCINDGSEKPKICTCKKCCAGKWIQLKQDEWENIKNRF
EQYKNGEGDIPVKIIEELISRLAAATKAERGLSLDKLITSLGCNDSQKKNTE
RDIVCLEKLOVKATCSQSHGDKTQCOQYETPEDEEDLLLEENENQVENQKV
NKAPFCPEKPEKPEKDBKQEBEKEQEBEKDKEGEEQSGAPSPSEVQETKL
PPPEPEEAKPPPVKPTAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
LGSATFLAKKKKSTIDARLNIPIKSDYDIPKLSPNRYIPYTSKYGKRYIYLR
GDSDTSGYTDHSDIISSESEYEEMDINDIYVPGSPKYLILIEVLEFPNGNNYLR
INTASQDNDIQNGIPSKNSDNEWNTLKDDFISQYLOEQPNLDPDKSGNVT
LNTEPTLYFDKPEKPFITSIHDRNLYTGEYSYNNMSTNSMDDPIINRDNVYSG
IDILNALNGDYDIYDEMLAKENELFCTNHVKQTSIHSVAKNTNSDPIHNOLELPHK
WLDHRDMCEQWENHERLAKLKEEWENETHSGNTHPSDSNKTINTDVSQIDMDGK
PKKEFTNMDTILLEDKYNERYVDQDDIYDVNDHDTSTVNNMKEPKSKVKTLDV
NKKTIKKEYPISDWDI"
/complement(29233..29304)
/notes="PFE0010c"
/notes="synonym: pseudo-VAR fragment"
/pseudo
/complement(29233..29304)
/notes="PFE0010c"
/notes="VAR fragment"
/pseudo
/codon_start=1
/complement(29923..30963)
/notes="RIF, degenerate"
/notes="Similar to Plasmodium falciparum rifin pfb1040W
TR:Q96293 (EMBL:AE001433) (345 aa) fasta scores: E():
1.1e-50, 53.161% id in 348 aa. Contains nonsense mutations
at apparent start position for exon 2"
/pseudo
/codon_start=1
/complement(30904..30909)
/notes="RIF, degenerate"
/notes="Possible splice donor"
/pseudo
/psuedo
complement(join(33056..34126,34325..34378))
/notes="RIP"
/notes="synonym: PFE0020c"
complement(join(33056..34126,34325..34378))
/notes="RIF"
/notes="Similar to Plasmodium falciparum rifin pfb0030C
TR:Q96112 (EMBL:AE001367) (370 aa) fasta scores: E():
3.3e-62, 53.125% id in 384 aa
1. probable transmembrane helix predicted for by TMHMM2.0
at aa 332-354"
/codon_start=1
/product="RIFIN"
/protein_id="CAD51370.1"
/db_xref="GI:23504492"
/db_xref="SPTREMBL:Q81494"
/translation="MKFSYFNILLFSLPLNTLINDHSKYSCKHTSNTKTPHRSLY
ECGLYSPANDNDPEMKRVQQDFDRTSRFHEYDERMQSRMCKQKQCKEIQKIL
KDLKQEQOQLNTLETKIDDDIPTCVCEKSLADKVEKGLRGCGYGLGTVAFTVGLI
GAVVHVWKMALAEATAAIAKASAAEISAANAAGIQAGKIAVESLKLIVYDFWP
EMSNVILNMHSHYGVANLTAFLHPKENVCKDAGEVILDKCNADFMDGFGILKKGVTN
GLLPKDAVPVLKGIQVCAQSPAKVADAARQVTYTAITEKETAINTIFMSKQTALL
ASVVAIVIVLMIILIIILIRKRRKKKKKKLQIILKULKE"
/complement(join(36464..37324,37496..37576))
/notes="RIF"
/notes="synonym: PFE0025c"
complement(join(36464..37324,37496..37576))
/notes="RIP"
/notes="Similar to Plasmodium falciparum rifin pfb1005W
TR:Q96288 (EMBL:AE001432) (316 aa) fasta scores: E():
8.3e-52, 54.574% id in 317 aa
2. probable transmembrane helices predicted for by TMHMM2.0
at aa 122-144 and 271-293"
/codon_start=1
/product="RIFIN"
/protein_id="CAD51371.1"
/db_xref="GI:23504493"
/db_xref="SPTREMBL:Q81493"
/translation="MAINMLKHSKILLFSLIOLNILLTSSVNNKKNPYLTPHTSAIT
SVLNEDCLYMPNYDKDADKSVKENDFDROSSRRPSEYVNNMTRNOKYKEQCKDQ
KIILKIEKLEKEKVEKCLICGGLGGVAPVFLGGLVNNMKAAVAADTVG
IKKAEKVGSIFFYLDAGSTWETKMINAGYSQKSLVEITILKNKEDEALAGPL
KCASEAATSGSEVFEFSGNISMAANAAGAAAGNAANGKAYEATSVGTICSNPVVISA
IVVVIATVILLIILIRYRRKKMKQKQYVTKLLNQ"
/complement(39443..40488)
/notes="STSVOR, degenerate"
/notes="synonym: PFE0030c"
/pseudo
complement(39443..40488)
/notes="STEVOR, degenerate"
/notes="Apparent single frameshift in exon II"
/pseudo
/codon_start=1
/complement(41129..41403)
/notes="RIF, degenerate"
/notes="synonym: PFE0035c"
/pseudo
complement(41129..41403)
/notes="RIF, degenerate"
/notes="RIF fragment"
/pseudo
/codon_start=1
/complement(join(43327..47481,47612..47761))
/notes="MESA"
/notes="synonym: PFE0040c"
complement(join(43327..47481,47612..47761))
/notes="MESA"
/notes="Similar to Plasmodium falciparum mature
parasite-infected erythrocyte surface antigen TR:Q06166
(EMBL:AF056936) (1661 aa) fasta scores: E(): 3.6e-144,
81.331% id in 1698 aa
Identical to pfemp2 (EMBL:AF056936) except in internal
repeat structures"
```

Query Match		3.3%; Score 70.4; DB 3; Length 347050;
Best Local Similarity		45.0%; Pred. No. 0.0071;
Matches 380; Conservative		0; Mismatches 461; Indels 3; Gaps 3;
		/codon_start=1
QY	1009	AAATTGGAAGAGTCTCTCTCCGCACTAACTCTTACTTACTCTAGCTATGTTAA 1068
Db	105990	ATATATATATACATT 106049
QY	1069	ATATTGAATGTTGGCACTCTCGTATTAAATATGCCAGTTGCACTAGATAAAAAACATG 1128
Db	106050	AAATTTCAATTCAGCAAAATATATAATATATATATATATATATATATATATAT 106109
QY	1129	ATAGACATTTAGTTTAAACTTGAATGTTATTTGAACTCTTTGGATACCGTGGATGTT 1188
Db	106110	TAATATAATAAGTAGTTAGTCATATATATATATATATATATATATATATATAT 106168
QY	1189	GTATGATTAATTTGAAGATATTTATATATTTGAAGATGTTTATATATATATAGATT 1248
Db	106169	TTATGTCAAATAAATTTGGCATATTCATTAATAATTAATTCATATATATATAAT 106228
QY	1249	TATAGAGAAAATTAATGATGATGCTGTGCTTTGTAGTTACTCTTTTTTGTTCGTA 1308
Db	106229	TATAACATTTCTTATATTTGTATAT-ATATATATATATATATTTTTTTTTTTTAT 106287
QY	1309	GTCCTTTCTCCATCCTCCTATGAAGAGAAAATCCAAATAGTTTAAGGAAAATTTTGT 1368
Db	106288	ATTATTTCTTTCAATGAAATAATAATGAAACTAAATTAATTAATTAAGTTTATTA 106347
QY	1369	AATTCATAGCTTTTTTCGTAAACACAGTCTATGTAGCTATCGTCATCATATTCCTTT 1428
Db	106348	TATTTTTTTAAATTTATATGAAATGTTATTTTATATATTTCCCTATATTTTTTTT 106407
QY	1429	GCAACAAAAAAGATCGTTTTTGTAAAAATTTAGTAGGCACATAAGTCGTCATTGTT 1488
Db	106408	ATGTTATCAATCTCGCTCAATTTGTTTTAAATTTATTCATTTATATGTTATTAAT 106467
QY	1489	GTCTCTCGAAATCTAGCGTTCTGTGCATCCACAAATAAGTTGTTGATTCGAGCTCCAA 1548
Db	106468	TATCTAATTTATTTATATATATTTTTTTTTTTTAAATTTATTTGTTAAATTTTTT 106527
QY	1549	GATTATAAATCTTTTATAGTGGTCAATGAAGATTTCTTAACCTCGTATACGAGTGATP 1608
Db	106528	TTAATTTATATAAATAATATATAT-ATAATATATATATATATATATATATATATA 106586
QY	1609	TATAATTTCPAACATATACGCTTTGTTTTTGGTAGGCTCTGCGCTTTTGAGACCACCC 1668
Db	106587	TTTATATTTATATATATATATTTTTTTTTTTTCTTTTCTTTCACGTTTGCATAATA 106646
QY	1669	CTTGCTAAATGTTTGTGGACCTTAGACAATCCATAATCGTTACGTTACGTCGAGTTGC 1728
Db	106647	TTTTTAATTAATTTAAGATATATAATAAGAAAAATATAAATAAATAACCAAAAAATA 106706
QY	1729	ACCAAAATGTCGCAATATAATTTAAATTTGGCCACAAAAACAATTTTCAACAAATAAT 1788
Db	106707	ATATAAATAATCCATTTAATAATAAACAATAATAATAATAATAAATAATCCATTAT 106766
QY	1789	CAACAACATCGATCGTTTCAAAATTTTATTTATTCATGCGGTTATTTGTTCAATGAAA 1848
Db	106767	AAATAACAAATAAAATATCATATATATATATATATATATATATATATATATATAT 106826
QY	1849	TATTT 1852
Db	106827	AGTT 106830
RESULT 15		
ATT32A17/c		
LOCUS		104386 bp DNA linear PLN 22-MAR-2000
DEFINITION		Arabidopsis thaliana DNA chromosome 4, BAC clone T32A17 (ESSA project).
ACCESSION		AL161813 AC007635
VERSION		AL161813.1 GI:7321058
KEYWORDS		Arabidopsis thaliana (thale cress)
SOURCE		Arabidopsis thaliana
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE		1 (bases 1 to 96453)
AUTHORS		McCombie,R.W., Siegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M., Vil,D.M., Preston,R.R., Matero,A., Shah,R., See,L.H., O'Shaughnessy,A., Rodriguez,M., Shekher,M., Schutz,K., Mayer,K.F.X., Swaby,I., Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and
JOURNAL		Unpublished
AUTHORS		2 (bases 95732 to 104386)
JOURNAL		McCombie,R.W., Robben,J., Grymonprez,B., Bastiaens,I., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
AUTHORS		Unpublished
JOURNAL		3 (bases 1 to 104386)
AUTHORS		EU Arabidopsis sequencing,project.
JOURNAL		Direct Submission
TITLE		Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
COMMENT		On Mar 23, 2001 this sequence version replaced gi:5103797. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES		Location/Qualifiers
source		1..104386
		/organism="Arabidopsis thaliana"
		/mol_type="genomic DNA"
		/variety="Columbia"
		/db_xref="taxon:3702"
		/chromosome="4"
misc_feature		1..22908
		/note="overlap to BAC T3F12; please refer to EMBL:AC002983 or analysis and annotation"
gene		22865..24293
		/gene="AT4G08700"
gene		complement(join(22865..23096,23227..24293))
		/genes="AT4G08700"
CDS		complement(join(22865..23096,23227..24293))
		/genes="AT4G08700"
		/note="similarity to various hypothetical proteins of Arabidopsis thaliana"
		Contains Prokaryotic membrane lipoprotein lipid attachment site AA82-92;Prokaryotic membrane lipoprotein lipid attachment site AA225-235
		contains EST gb:A1997220.1"
		/codon_start=1
		/product="putative protein"
		/protein_id="CA882106.1"
		/db_xref="GI:7321059"
		/translation="MLLKEDEGTRTSVPTQIMKLRTHWILVIFISIAQA IAVLLGRFYVNGGSKWISTLVQTCGFPILYPLCLFPASHSSSSCSKPTIVWYLSUGFAIGLNLVYSGLLVLSASTYSITGLSSOLAFNGVFSYINSQKITCLILFVLFLSVAVILSDDDNSPFGSKWYLGICLCTVFASLIYSLQSLQFSPENLVKSETFSWLEMQIYLSLVASCVAIVGFASGEWMLLSVEMEHEGQVLYULVGTAVSQWLGSGVAIVLIFVSSLSFNLIGTSLIVTLPAAIVFHDKLIVTKVWVAMLIAPMGQFVIYQNYLDDLKVQRAEAQAAYFLRAKVLCTIYIRNSTYEVKNGGFIENLGDTHV VSEDHVDADLETNVSFNKHVDSTLKASSSKRMVKIGF"
		complement(22865..23096)
exon		/gene="AT4G08700"
		/number=1
intron		complement(23097..23226)
		/gene="AT4G08700"
		/number=1
exon		complement(23227..24293)
		/gene="AT4G08700"
		/number=2
gene		24970..27800

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 02:09:18 ; Search time 859 Seconds
(without alignments)
10627.913 Million cell updates/sec

Title: US-10-027-880-1
Perfect score: 2149
Sequence: 1 gtcgaattggtatattgt.....actaatgctaatatcaag 2149

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002s.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2149	100.0	2149	4	Aaf25316 Nucleotid
2	2149	100.0	4280	4	Aaf25317 Nucleotid
3	2149	100.0	4309	4	Aaf25319 Nucleotid
4	2149	100.0	4413	4	Aaf25318 Nucleotid
5	63.6	3.0	7008	4	Aas46532 Tumour su
6	60.2	2.8	6175	6	ABL33307
7	59.8	2.8	7597	6	ABL33013 Human imm
8	59.6	2.8	8822	6	ABL33097 Human imm
9	59.4	2.8	8920	2	AAQ62924 Carbamoyl
10	59.4	2.8	33053	6	ABQ67005 Human ang
11	58.6	2.7	5413	4	AAS46694 Tumour su
12	58.2	2.7	6045	6	ABK31541 Signal tr
13	58.2	2.7	6045	6	ABL70624 Chemicall
14	58.2	2.7	8056	6	ABL70100 Haematopo
15	57.6	2.7	37515	6	ABQ66998 Human ang
16	57.2	2.7	8056	6	ABZ10246 Haematopo
17	56.8	2.6	8277	6	ABL33328 Human imm
18	56.6	2.6	6292	4	AAS46735 Tumour su
19	56.6	2.6	8056	7	ABL10246 Haematopo
20	56.4	2.6	15951	6	ABL33680 Human imm
21	56.4	2.6	15951	6	ABL34580 Human met
22	56.4	2.6	15951	6	ABL70373 Chemicall
23	56.2	2.6	8056	7	ABZ10100 Haematopo

C	24	56.2	2.6	17280	4	AAS46772 Tumour su
	25	56	2.6	37973	6	ABL34197 Human imm
	26	55.8	2.6	6076	4	AAS46664 Tumour su
C	27	55.6	2.6	393	7	ABX39417 Bovine ES
	28	55.6	2.6	7072	6	ABK31471 Signal tr
	29	55.6	2.6	7072	6	ABL70566 Chemicall
	30	55.6	2.6	7072	6	AAS61385 Human gen
	31	55.6	2.6	7922	6	ABN80086 Human che
	32	55.4	2.6	12327	6	ABL34358 Human imm
	33	55.4	2.6	15224	6	ABK31358 Signal tr
	34	55.4	2.6	15224	6	ABL70317 Chemicall
	35	55.4	2.6	15224	6	AAS61262 Human gen
	36	55.4	2.6	110000	7	AAAD53224_2
	37	55	2.6	11745	6	ABK28332 DNA trans
C	38	55	2.6	113515	6	ABL34174 Human imm
	39	54.8	2.6	5297	4	AAS46513 Tumour su
	40	54.8	2.6	14919	4	AAS46506 Tumour su
	41	54.6	2.5	8244	4	AAS46395 Tumour su
	42	54.6	2.5	8244	6	ABL32992 Human imm
	43	54.6	2.5	8244	6	ABQ67031 Human arg
	44	54.6	2.5	8244	9	ADB54133 Pretreat
	45	54.6	2.5	13131	6	ABL92248 Chemicall

ALIGNMENTS

RESULT 1
ID AAF25316 standard; DNA; 2149 BP.
XX
AC AAF25316;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant promoter for expression in roots.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX Arabidopsis thaliana.
OS
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR001768.
XX
PR 25-JUN-1999; 99FR-00008185.
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PI Hoffmann B, Mollier P, Pelletier G;
XX WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 1; Fig 1; 92pp; French.
XX
CC The present sequence represents a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 2149 BP; 693 A; 365 C; 335 G; 756 T; 0 U; 0 Other;

Query Match	100.0%;	Score 2149;	DB 4;	Length 2149;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 2149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	GTGCAATGCGATATATCTGTAAGCAATCTGAAAGAAATAGTGGGATATATAAACAACCG	60		
DB	1	GTGCAATGCGATATATGTAAGCAATCTGAAAGAAATAGTGGGATATATAAACAACCG	60		
QY	61	GCGAAGTCAAGTCTCTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT	120		
DB	61	GCGAAGTCAAGTCTCTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT	120		
QY	121	AATTCCTGAATCTTTCATCTTCGAAATGATATTTATCAATTTTATCAAAAAAAGTACA	180		
DB	121	AATTCCTGAATCTTTCATCTTCGAAATGATATTTATCAATTTTATCAAAAAAAGTACA	180		
QY	181	AGTTCTACCAAGCAGAGGATTTAAACAACTTGTGTGCAAAATGCTAAATTTAAAGCCTAA	240		
DB	181	AGTTCTACCAAGCAGAGGATTTAAACAACTTGTGTGCAAAATGCTAAATTTAAAGCCTAA	240		
QY	241	TCTTATGATTTCCCTTTCTTTCAGGATATATCTGATATTTGATATGACCCATTTGTTTG	300		
DB	241	TCTTATGATTTCCCTTTCTTTCAGGATATATCTGATATTTGATATGACCCATTTGTTTG	300		
QY	301	TCATTAACCTCCACCTCTATACATCAGTACTCAGAGTCGAATAACAATATCCATAAGAA	360		
DB	301	TCATTAACCTCCACCTCTATACATCAGTACTCAGAGTCGAATAACAATATCCATAAGAA	360		
QY	361	GTGGTATATTGTGAAAAAATAAAAAAAGTGGTATATCTGGTATATACAATACACGGTC	420		
DB	361	GTGGTATATTGTGAAAAAATAAAAAAAGTGGTATATCTGGTATATACAATACACGGTC	420		
QY	421	TCGAAATGCGCTCAACAAATTTCTAGGAGAAATGGAAGTGTCTCTTTGGTTTATTTATT	480		
DB	421	TCGAAATGCGCTCAACAAATTTCTAGGAGAAATGGAAGTGTCTCTTTGGTTTATTTATT	480		
QY	481	CTTAATAACATACTCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTCGAATGTCC	540		
DB	481	CTTAATAACATACTCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTCGAATGTCC	540		
QY	541	TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTGCAAGCCACAGGACATGCATAG	600		
DB	541	TAAATTTCTCTAATCATAAATCGTAAAGAAAATTCGTGCAAGCCACAGGACATGCATAG	600		
QY	601	GGCAGTAGTACCTTTTAAACCAATCAAAATATATTAAAGAAAAGAACTTCCTTAAA	660		
DB	601	GGCAGTAGTACCTTTTAAACCAATCAAAATATATTAAAGAAAAGAACTTCCTTAAA	660		
QY	661	AGAACAAATTAATAAGTGGATAAAAAGATAAGAGGTAGGACGAAGAAAACGTATGG	720		
DB	661	AGAACAAATTAATAAGTGGATAAAAAGATAAGAGGTAGGACGAAGAAAACGTATGG	720		
QY	721	CCGGACTCGTAAACAGGGAGCTCCGACCACTCCGAGACGGCGACGCTGACTGAT	780		
DB	721	CCGGACTCGTAAACAGGGAGCTCCGACCACTCCGAGACGGCGACGCTGACTGAT	780		
QY	781	TTTTCTTTTCTTTTCTTAAAGACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC	840		
DB	781	TTTTCTTTTCTTTTCTTAAAGACGTTGTTTCGTGCTTACAAGGGTCAAAACCATATCC	840		
QY	841	AATTTGTTCTGCTATTATATATAACTAAAGATCCCTCTTGCTGCTTTGCTTTTATTCGT	900		
DB	841	AATTTGTTCTGCTATTATATATAACTAAAGATCCCTCTTGCTGCTTTGCTTTTATTCGT	900		
QY	901	GATATATAACTTAACCTTAAATAGTCTTAAATATATATGTCCTTACTGTTTCTACTG	960		
DB	901	GATATATAACTTAACCTTAAATAGTCTTAAATATATATGTCCTTACTGTTTCTACTG	960		
QY	961	ACCTCAGTCCCTAGTTAGCTATATGGAATATGGAATATGGAATATGGAATATGGAAT	1020		
DB	961	ACCTCAGTCCCTAGTTAGCTATATGGAATATGGAATATGGAATATGGAATATGGAAT	1020		
QY	1021	TTCCCTCTCTGCAACTAACTCTTATCTTACTTACCTAGCTATGTTAAATATTTGAATGTT	1080		

Db	2101	TAAAGAGCTCTTATATCAACCACTTATCCCACTAAATGCTAAATATACAG	2149
RESULT 2			
AAF25317			
ID	AAF25317	standard; DNA; 4280 BP.	
XX			
AC	AAF25317;		
XX			
DT	30-APR-2001	(first entry)	
XX			
DE	Nucleotide sequence of a root promoter and the GSU gene coding region.		
XX			
KW	Plant promoter; root cell; root-specific expression; parasite resistance;		
KW	nematode resistance; fungal resistance; water stress; salt stress;		
KW	sugar content; nitrogen transport; GUS gene; ss.		
XX			
OS	Synthetic.		
OS	Arabidopsis thaliana.		
OS	Unidentified.		
XX			
FN	WO200100833-A1.		
XX			
PD	04-JAN-2001.		
XX			
XX	23-JUN-2000;	2000WO-FR001768.	
PF			
XX			
PR	25-JUN-1999;	99FR-00008185.	
XX			
PA	(INRG) INRA INST NAT RECH AGRONOMIQUE.		
XX			
PI	Hoffmann B, Mollier P, Pelletier G;		
XX			
DR	WPI; 2001-102893/11.		
XX			
PT	New constitutive plant promoter active specifically in roots, useful for		
PT	controlling expression of pest or drought resistance genes, and related		
PT	transgenic plants.		
XX			
PS	Claim 5; Page 69-71; 92pp; French.		
XX			
CC	The present sequence represents a plant promoter linked to a GUS coding		
CC	region. The plant promoter directs expression of a selected sequence in		
CC	root cells at all stages of development of a plant. The plant promoter is		
CC	used to control expression of genes in a root-specific manner, especially		
CC	genes that provide resistance to parasites, pests (nematodes or fungi),		
CC	water and salt stress, or alter sugar content or nitrogen transport.		
CC	Fragments of the promoter are useful as probes or primers to detect or		
CC	amplify at least part of the promoter		
XX			
SQ	Sequence 4280 BP; 1256 A; 859 C; 907 G; 1258 T; 0 U; 0 Other;		
	Query Match	100.0%; Score 2149; DB 4; Length 4280;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 2149; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	GTGGAATTGTGATATATTGTAAGCAATCTGAAAGAAATAGTGGATATATAACAACCG	60
Db	1	GTGGAATTGTGATATATTGTAAGCAATCTGAAAGAAATAGTGGATATATAACAACCG	60
QY	61	GGGAAGTACAAGTCTACCTTTTGGCATGGAAACCATGTTTGTAGGATTTACTTTGT	120
Db	61	GGGAAGTACAAGTCTACCTTTTGGCATGGAAACCATGTTTGTAGGATTTACTTTGT	120
QY	121	AATTCTCGAATCTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACA	180
Db	121	AATTCTCGAATCTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACA	180
QY	181	AGTTCTACAAAGCAGAGGTTAAACAACCTGTGTCAAATGCTAATTTAAAGCCTAA	240
Db	181	AGTTCTACAAAGCAGAGGTTAAACAACCTGTGTCAAATGCTAATTTAAAGCCTAA	240
QY	241	TCTTATGATTTCCCTTTTCTTCACGATATATCTGATATGATATGACCACTTTGTTG	300

Db 1321 CATCTCTCTATGAGAAAAATCAAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1380
Qy 1381 TTTTTCGTAACCAAGTCTCTATGATAGTATCGTTCATCATATTCCTCTTTGCAACAACAAAA 1440
Db 1381 TTTTTCGTAACCAAGTCTCTATGATAGTATCGTTCATCATATTCCTCTTTGCAACAACAAAA 1440
Qy 1441 AGATCGTTTTGTAAAAATTTAGTAGGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Db 1441 AGATCGTTTTGTAAAAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTCGAAA 1500
Qy 1501 TCTAGCGTTCTGTCATCCCAATAAGTTGTTGATTTCGAGCTTCAAGATTAATAATCTT 1560
Db 1501 TCTAGCGTTCTGTCATCCCAATAAGTTGTTGATTTCGAGCTTCAAGATTAATAATCTT 1560
Qy 1561 TTTTAGATGGTTCATGAAGATTTCTAACTTCGATACGAGTGTATCCATATAATTTCTAA 1620
Db 1561 TTTTAGATGGTTCATGAAGATTTCTAACTTCGATACGAGTGTATCCATATAATTTCTAA 1620
Qy 1621 CATATACGTTCTGTTTGGTAGGCTCTGCGTCTTTTGAGACCAACCCCTTGCTAAATGTT 1680
Db 1621 CATATACGTTCTGTTTGGTAGGCTCTGCGTCTTTTGAGACCAACCCCTTGCTAAATGTT 1680
Qy 1681 TTCTGTCACCTTAGACAATCCATAACGTTACGTCGAGTTCGAGTTCGACCAAAATGGTC 1740
Db 1681 TTCTGTCACCTTAGACAATCCATAACGTTACGTCGAGTTCGAGTTCGACCAAAATGGTC 1740
Qy 1741 CAAATATAATTTAAATTTGGCCCAAAAACAACATTTTACAAAACAAATTCACAAACATGC 1800
Db 1741 CAAATATAATTTAAATTTGGCCCAAAAACAACATTTTACAAAACAAATTCACAAACATGC 1800
Qy 1801 ATCGTTTCAAAATTTATTTTCAATGCGCTATTTGTTTCAATGTTTCAATATTCGTTTAA 1860
Db 1801 ATCGTTTCAAAATTTATTTTCAATGCGCTATTTGTTTCAATGTTTCAATATTCGTTTAA 1860
Qy 1861 CTCACATGACCAATTTTAAATTTTCAAGAAGAACATTTTGTATATAAATAACATTTTA 1920
Db 1861 CTCACATGACCAATTTTAAATTTTCAAGAAGAACATTTTGTATATAAATAACATTTTA 1920
Qy 1921 TGAACCAACCGGTAAAGTCGATGATTTTCAAGTTTATGTTTGTGTTTGTGTAATCAT 1980
Db 1921 TGAACCAACCGGTAAAGTCGATGATTTTCAAGTTTATGTTTGTGTTTGTGTAATCAT 1980
Qy 1981 TAAGGACTACATTTGATCCCTCATCTTAAATAATAGGAATCAACATGATGATTA 2040
Db 1981 TAAGGACTACATTTGATCCCTCATCTTAAATAATAGGAATCAACATGATGATTA 2040
Qy 2041 GTTCACCAAAAGAGCTCTCTATGCGCTAATTAAGAGTCAGACGAGGATGACCGGGTCAT 2100
Db 2041 GTTCACCAAAAGAGCTCTCTATGCGCTAATTAAGAGTCAGACGAGGATGACCGGGTCAT 2100
Qy 2101 TAAGAGCTCTATATTCACCAATTTCTCCACTAATTCGCTAATTAATCAG 2149
Db 2101 TAAGAGCTCTATATTCACCAATTTCTCCACTAATTCGCTAATTAATCAG 2149

RESULT 3

AAF25319 standard; DNA; 4309 BP.

AAF25319;

30-APR-2001 (first entry)

Nucleotide sequence of a plant promoter insert in plasmid pbin19.

Plant promoter; root cell; root-specific expression; parasite resistance;

nematode resistance; fungal resistance; water stress; salt stress;

sugar content; nitrogen transport; ss.

Synthetic.

Arabidopsis thaliana.

Unidentified.

Key Location/Qualifiers
promoter 30..2178
misc_feature /tag= a
/tag= b
/note= "GUS coding region"
WO200100833-A1.
04-JAN-2001.
23-JUN-2000; 2000WO-FR001768.
25-JUN-1999; 99FR-00008185.
(INRG) INRA INST NAT RECH AGRONOMEQUE.
Hoffmann B, Mollier P, Pelletier G;
WPI; 2001-102893/11.
New constitutive plant promoter active specifically in roots, useful for
controlling expression of pest or drought resistance genes, and related
transgenic plants.
Example 3; Page 73-75; 92pp; French.
The present sequence represents a plant promoter insert in plasmid
pbin19. The plant promoter directs expression of a selected sequence in
root cells at all stages of development of a plant. The plant promoter is
used to control expression of genes in a root-specific manner, especially
genes that provide resistance to parasites, pests (nematodes or fungi),
water and salt stress, or alter sugar content or nitrogen transport.
Fragments of the promoter are useful as probes or primers to detect or
amplify at least part of the promoter
SQ Sequence 4309 BP; 1262 A; 868 C; 915 G; 1264 T; 0 U; 0 Other;

Query Match 100.0%; Score 2149; DB 4; Length 4309;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGAATTTGATATATTGTAAGCAATCTGAAAGATAAGTGGGATATATAACAACCG 60
Db 30 GTCGAATTTGATATATTGTAAGCAATCTGAAAGATAAGTGGGATATATAACAACCG 89
Qy 61 GCBAAGTACAAGTTCTACCTTTTGGCATGGAACCACTTTTAGGATTTACTTTGT 120
Db 90 GCBAAGTACAAGTTCTACCTTTTGGCATGGAACCACTTTTAGGATTTACTTTGT 149
Qy 121 AATTCTGGAATCTTTCATTTCTTGAAATTTGATATTTACATTTTATCAAAAAAAGTACA 180
Db 150 AATTCTGGAATCTTTCATTTCTTGAAATTTGATATTTACATTTTATCAAAAAAAGTACA 209
Qy 181 AGTTCTACCAAGCAGGAGTTAAACAACCTTGTGTCAAACTGCTAAATTTAAAGCCTAA 240
Db 210 AGTTCTACCAAGCAGGAGTTAAACAACCTTGTGTCAAACTGCTAAATTTAAAGCCTAA 269
Qy 241 TCTTATGATTTCCCTTTCTTCCAGGATATATCTGATTTGATATGCAACCATTTGTTG 300
Db 270 TCTTATGATTTCCCTTTCTTCCAGGATATATCTGATTTGATATGCAACCATTTGTTG 329
Qy 301 TCATTAATCTCCCACTCTATACATCAGTATCTCAAACTGCAATTAACAATATCCATAAGAA 360
Db 330 TCATTAATCTCCCACTCTATACATCAGTATCTCAAACTGCAATTAACAATATCCATAAGAA 389
Qy 361 GTGCTATATTTGTAAGAAAAAAGGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 420
Db 390 GTGCTATATTTGTAAGAAAAAAGGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 449
Qy 421 TCGAATTTGCTTCACCAATTTCTAGGAGAAATGACGCTGCTCTTTGTTGTTTATTATT 480
Db 450 TCGAATTTGCTTCACCAATTTCTAGGAGAAATGACGCTGCTCTTTGTTGTTTATTATT 509

```
QY 481 CTTAATAACACACCTCTATATTTTAAACACACTCGATGCTCGCTTAAATTTTCAATGTGCC 540
DB 510 CTTAATAACACTCTATATTTTAAACACACTCGATGCTCGCTTAAATTTTCAATGTGCC 569
QY 541 TAAATTTCTCTAATCATAAATCGTAAGAAATTCGTGGAAGCCACAGGACATGCATAG 600
DB 570 TAAATTTCTCTAATCATAAATCGTAAGAAATTCGTGGAAGCCACAGGACATGCATAG 629
QY 601 GGCACTGATGTACCTTTAAAAACCATCAAAAAATATTAATAGAAAAGGAAATTCCTTAAA 660
DB 630 GGCACTGATGTACCTTTAAACCATCAAAAATATTAATAGAAAAGGAAATTCCTTAAA 689
QY 661 AGAACATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGAGAGAAAAAGTATGG 720
DB 690 AGAACATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGAGAGAAAAAGTATGG 749
QY 721 CCGGACTCGTAACAAGGAGCTCCGACCACTCGGAGAGCGGAGACGCTGACTGATT 780
DB 750 CCGGACTCGTAACAAGGAGCTCCGACCACTCGGAGAGCGGAGACGCTGACTGATT 809
QY 781 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTTACAAGGGTCAAAACCATATCC 840
DB 810 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTTACAAGGGTCAAAACCATATCC 869
QY 841 AATGTTCTGCTATATATATACTAAAGATCCCTCTTGTCTTTGCTTTTATTCGT 900
DB 870 AATGTTCTGCTATATATATACTAAAGATCCCTCTTGTCTTTGCTTTTATTCGT 929
QY 901 GATATATACTAACTAAATAGTCTTAAATATATATATGTCCTACCTAGTGTTCCTACTG 960
DB 930 GATATATACTAACTAAATAGTCTTAAATATATATATGTCCTACCTAGTGTTCCTACTG 989
QY 961 ACCTCAGTCCCTAGTATATATGACATATGTGAAATGACGCCCAAAATTTGAAGAG 1020
DB 990 ACCTCAGTCCCTAGTATATGACATATGTGAAATGACGCCCAAAATTTGAAGAG 1049
QY 1021 TTCTCTCTCTGCAACTTAATCTTATCTTACTCAATGAGCTATGTTAAATATTGAATGTT 1080
DB 1050 TTCTCTCTCTGCAACTTAATCTTATCTTACTCAATGAGCTATGTTAAATATTGAATGTT 1109
QY 1081 GGCATCTCTGATTAATAATGCCAGTTGCCACCTAGATAAAAAACATGATAGACATTTAG 1140
DB 1110 GGCATCTCTGATTAATAATGCCAGTTGCCACCTAGATAAAAAACATGATAGACATTTAG 1169
QY 1141 TTTAAAACTTGAATGTTATTTGAACCTCTTTGGATTAAGTGGATTGTTGATGGAATAAA 1200
DB 1170 TTTAAAACTTGAATGTTATTTGAACCTCTTTGGATTAAGTGGATTGTTGATGGAATAAA 1229
QY 1201 TTTTGAAGATTTATATATTTGAAGATGTTTATATATATATATAGATTTTATATAGCAAAAA 1260
DB 1230 TTTTGAAGATTTATATATTTGAAGATGTTTATATATATATATAGATTTTATATAGCAAAAA 1289
QY 1261 TATTGATGATAGTCTCTCTTTGTTAGTACTCTTTTGTGTCGGTAGTCCCTTCTCCT 1320
DB 1290 TATTGATGATAGTCTCTCTTTGTTAGTACTCTTTTGTGTCGGTAGTCCCTTCTCCT 1349
QY 1321 CATCCTCCTATGAAGAAAAATCCAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1380
DB 1350 CATCCTCCTATGAAGAAAAATCCAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1409
QY 1381 TTTTCGTAACACAGTCTTATGATGATCTATCATCATATATTCCTCTTTTGAACAACAAAA 1440
DB 1410 TTTTCGTAACACAGTCTTATGATGATCTATCATCATATATTCCTCTTTTGAACAACAAAA 1469
QY 1441 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAGTCGTCAATTTGTTGCTGTCGAAA 1500
DB 1470 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAGTCGTCAATTTGTTGCTGTCGAAA 1529
QY 1501 TCTAGCGTTCTGTCATCCACAAAATAGTTGTTTCTGATTCGAGCTTCCAAGATTTAATCTTT 1560
DB 1530 TCTAGCGTTCTGTCATCCACAAAATAGTTGTTTCTGATTCGAGCTTCCAAGATTTAATCTTT 1589
```

```
QY 1561 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATAAATTTCTAA 1620
DB 1590 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATAAATTTCTAA 1649
QY 1621 CATATAGTCTTGTGTTTTTGGTAGGCTCTGGCTCTTTTGGAGACACCCCTTGGCTAATGTT 1680
DB 1650 CATATAGTCTTGTGTTTTTGGTAGGCTCTGGCTCTTTTGGAGACACCCCTTGGCTAATGTT 1709
QY 1681 TTGTTGCACCTTAGACAAATCCATAATAGTTTACGTGAGTCGAAGTTGCACCAAAATGTC 1740
DB 1710 TTGTTGCACCTTAGACAAATCCATAATAGTTTACGTGAGTCGAAGTTGCACCAAAATGTC 1769
QY 1741 CAAATATAATTTAAATTTGGCCACAAACCAACATTTTACAAACAAATTCACAAACATGC 1800
DB 1770 CAAATATAATTTAAATTTGGCCACAAACCAACATTTTACAAACAAATTCACAAACATGC 1829
QY 1801 ATCGTTTTCAAAATTTTATTTTCAATGCGGTATTTGTTCAATTTGTTAAATATTTCTGTTAA 1860
DB 1830 ATCGTTTTCAAAATTTTATTTTCAATGCGGTATTTGTTCAATTTGTTAAATATTTCTGTTAA 1889
QY 1861 CTCACTGACGAATTTTAAATTTTCAAGAAAGAACATTTTGTATATAATAACATTTTA 1920
DB 1890 CTCACTGACGAATTTTAAATTTTCAAGAAAGAACATTTTGTATATAATAACATTTTA 1949
QY 1921 TGGAAACACCGGTTAAGCTCGATGATTTTGAAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 1980
DB 1950 TGGAAACACCGGTTAAGCTCGATGATTTTGAAGTTTTAGTTTTGTCGTTTTGTGAAATCAT 2009
QY 1981 TAAGCACTACATTTGATCCCTCATTTACTTTAATAATAGGAATCAAAACATGATGATTA 2040
DB 2010 TAAGCACTACATTTGATCCCTCATTTACTTTAATAATAGGAATCAAAACATGATGATTA 2069
QY 2041 GTTCACCAAAAGAGCTCTTTATGCTATTAAAGAGTCAGACGAGGATGACCGGGGTGAT 2100
DB 2070 GTTCACCAAAAGAGCTCTTTATGCTATTAAAGAGTCAGACGAGGATGACCGGGGTGAT 2129
QY 2101 TAAGAGCTCTTATATTTCAACCATTTACTCCACTAATTCCTAATTAATCAG 2149
DB 2130 TAAGAGCTCTTATATTTCAACCATTTACTCCACTAATTCCTAATTAATCAG 2178
```

RESULT 4

AAF25318
ID AAF25318 standard; DNA; 4413 BP.

XX AAF25318;

XX AC AC
XX XX
DT 30-APR-2001 (first entry)

XX Nucleotide sequence of an Arabidopsis thaliana genomic fragment.

DE DE
XX XX
KW KW
KW nematode resistance; root cell; root-specific expression; parasite resistance;
KW sugar content; nitrogen transport; ss.
XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX WO200100833-A1.

PN 04-JAN-2001.

XX 23-JUN-2000; 2000WO-FR001768.

XX 25-JUN-1999; 99FR-00008185.

XX (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX Hoffmann B, Mollier P, Pelletier G;

XX WPI; 2001-102893/11.

XX New constitutive plant promoter active specifically in roots, useful for

PT controlling expression of pest or drought resistance genes, and related

PT transgenic plants.

PS
XX
Claim 3; Page 71-73; 92pp; French.

CC The present sequence represents a genomic fragment of *Arabidopsis*
CC *thaliana*. The sequence contains a plant promoter that directs expression
CC of a selected sequence in root cells at all stages of development of a
CC plant. The plant promoter is used to control expression of genes in a
CC root-specific manner, especially genes that provide resistance to
CC parasites, pests (nematodes or fungi), water and salt stress, or alter
CC sugar content or nitrogen transport. Fragments of the promoter are useful
CC as probes or primers to detect or amplify at least part of the promoter

SQ Sequence 4413 BP; 1427 A; 731 C; 702 G; 1553 T; 0 U; 0 Other;

Query Match 100.0%; Score 2149; DB 4; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0;

Qy	1	GTCGAA	TGTCGATATATTTGTTAAGCAATCTGAAAAAGATAAGTGGGATATATAAACCAACCG	60
Db	136	GTCGAA	TTGTGATATATTTGTTAAGCAATCTGAAAAAGATAAGTGGGATATATAAACCAACCG	195
Qy	61	GCGAAAGT	ACAAAGTCTACCTTTTTTGGCATGGAAACCATGTTTTAGGATTTACTTTGT	120
Db	196	GCGAAAGT	ACAAAGTCTACCTTTTTTGGCATGGAAACCATGTTTTAGGATTTACTTTGT	255
Qy	121	AAATTCCT	GAAATCTTTCATTTCTTGAAATTTGATATTTACATTTTTATCAAAAAAAAAGTACA	180
Db	256	AAATTCCT	GAAATCTTTCATTTCTTGAAATTTGATATTTACATTTTTATCAAAAAAAAAGTACA	315
Qy	181	AGTTCTAC	CAAAAGCACAGGAGTTAAACAACTGTGTGTCAAATGCTAAATTTAAAGCCTAA	240
Db	316	AGTTCTAC	CAAAAGCACAGGAGTTAAACAACTGTGTGTCAAATGCTAAATTTAAAGCCTAA	375
Qy	241	TCTTAATG	ATTTCCCTTTTCTTCACGATATATATCTGATATTGATATGCAACCCATTTCTTTTG	300
Db	376	TCTTAATG	ATTTCCCTTTTCTTCACGATATATATCTGATATTGATATGCAACCCATTTCTTTTG	435
Qy	301	TCATTTAA	CTCCCTCATATACATCAGATATCTCAAAGTCGGAATTAACAATATCCATAAGAA	360
Db	436	TCATTTAA	CTCCCTCATATACATCAGATATCTCAAAGTCGGAATTAACAATATCCATAAGAA	495
Qy	361	GTGGTATAT	TGTGAAAAAAGTGGGTACTGTGATATACAAATACCAAGTCCACGGTC	420
Db	496	GTGGTATAT	TGTGAAAAAAGTGGGTACTGTGATATACAAATACCAAGTCCACGGTC	555
Qy	421	TCGAATTC	CGCTCAAAATTTCTAGGAGAAAATGGACGTCTCTCTTTGGTTTTATTTTATT	480
Db	556	TCGAATTC	CGCTCAAAATTTCTAGGAGAAAATGGACGTCTCTCTTTGGTTTTATTTTATT	615
Qy	481	CTTAATTA	ACATACCTATATTTTAAACACTTCGATGCTCTCGCTTAAATTTCGAATCTGCC	540
Db	616	CTTAATTA	ACATACCTATATTTTAAACACTTCGATGCTCTCGCTTAAATTTCGAATCTGCC	675
Qy	541	TAAATTTCT	CTAAATCATATAATCGTAAAGAAAATTCGTCGAAGCCACAGGAGCATGCTAG	600
Db	676	TAAATTTCT	CTAAATCATATAATCGTAAAGAAAATTCGTCGAAGCCACAGGAGCATGCTAG	735
Qy	601	GGCAGTAT	GTACCTTTAAACCCATCAAAAAATATATTAATAGAAAAAGGAAACTTCTTAA	660
Db	736	GGCAGTAT	GTACCTTTAAACCCATCAAAAAATATATTAATAGAAAAAGGAAACTTCTTAA	795
Qy	661	AGAACAA	TTTAATTAAGTGGATATAAAGATATAGAGGTAGGCGAAGAAACGATATGG	720
Db	796	AGAACAA	TTTAATTAAGTGGATATAAAGATATAGAGGTAGGCGAAGAAACGATATGG	855
Qy	721	CGCGACTC	GTAAACAAAGGACGCTCCCGACCACTGCGGAGACGCGAGACGCTGACTGATT	780
Db	856	CGCGACTC	GTAAACAAAGGACGCTCCCGACCACTGCGGAGACGCGAGACGCTGACTGATT	915
Qy	781	TTTTCTTT	TTCTTTCTTAAGAACGTTGTTTCTGTCCTTACAAAGGTCATAAACCATATCC	840


```
QY 1445 TCGTTTTGTAATAATTTAGTGGCAGTAAAGTCGTCATTTGTGCTCTCGAAATCTA 1504
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 530 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATG 589
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1505 GCGTCTGTCTATCCACAAATAAGTTGTTGATTCGAGCTTCCAGATTAATACCTTTTT 1564
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 590 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 649
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1565 AGATGGCTATGAAGATTCTCACTTCGTATACGAGTGATGCCATAAATTTCTAACATA 1624
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 650 TAGTGTGTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 705
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1625 TAGCTCTGTTTTGGTAGGCTCGCGCTTTGGAGACCCGCCCTTCGTAATGTTTGT 1684
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 706 TAGCTATTATATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 765
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1685 TGCACCTTAGACATCCATATAGCTTACGTGAGTCGAAGTTGCACCAAAATGTCCTCAA 1744
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 766 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 825
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1745 TATAAATTAAATTTGGCCACAAACAAATTTTCAAAACAAATTCACAAACATGCATCG 1804
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 826 ATTAATTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTC 885
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1805 TTTCAAAATTTATTATCAATGCGGTTATTTGTCATG-----TAAATATCTGTTTA 1859
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 886 GTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 945
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1860 ACTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTTCATATAAATTAACATTTT 1919
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 946 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1005
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1920 ATGGAACACCGGTTAAGCTCGATGATTTTGAAGTTTGTGCTTTGTGGAATCA 1979
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1006 ATGATTTATTATGATTTATGATTTATGATTTATGATTTATGATTTATGATTTATTA 1065
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1980 TTAACGACCTACATTTGATCCCTCATTTACTTTAATAATTAAGGAAT 2024
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1066 TTTATTATTATTAGTTAATAATTTATTATTATTATTATTATTATTATTATTATTAGTAT 1110
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 8
ABL33097
ID ABL33097 standard; DNA; 5822 BP.
XX
AC ABL33097;
XX
DT 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1070.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianemic; cytostatic; nontropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
```

```
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1070; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX SQ Sequence 5822 BP; 1880 A; 30 C; 1198 G; 2714 T; 0 U; 0 Other;
XX
XX Query Match 2.8%; Score 59.6; DB 6; Length 5822;
XX Best Local Similarity 51.4%; Pred. No. 0.045;
XX Matches 189; Conservative 0; Mismatches 179; Indels 2; Gaps 2;
QY 1091 TATTAAATATGCCAGTTGCCACCTAGATATAAAACATGATAGACATTAGTTTAAAACTT 1150
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4630 TTTTAAATATATATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 4689
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1151 GAAATGTTATTGACATCTTTGGATACGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1210
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4690 TTTATTTTAGTTTGTAGAAATTAATATATATGATTGATTGATTGATTGATTGATTGATT 4749
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1211 ATTATATATTGAGATGTTTATATATATATGATGTTTATATA-GCAGAAATATTGATGT 1269
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4750 GTTTTTTTAGTGATATTTTTTTTAAATTTTAGTTTTTATATAAATATTGTTTTTTT 4809
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1270 AGATGTTGCTCTTTTGTAGTTACTCTTTTTTGTGGCTAGTCTCTTCTCCTCATCCCT 1329
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4810 TGTAGGTTTGGGTGTTGTTTTTATGTTTTTATGTTGTTTATGTTTATGTTTATGTTT 4869
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1330 ATGAAGAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGCTCTTTTCGTAA 1389
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4870 TTTAGTGAATATTAGTGTGTTGTAGGAAATATTTTTTGGTTTTTTTAGGTTTTTTTTT 4929
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1390 CCACAGTTCTATGATCTATCGTCATCATATTCCTCTTTGCAACAACAAAAGATCGTT 1449
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4930 TGTGGATATTTTTTTAG-AAATGTTATTTTTGTTTTGTTTAAATGTTTAAATAGTTGTT 4988
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1450 TTTGTAAAAAT 1459
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 4989 TTTATTATTAAT 4998
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 9
AAQ62924
ID AAQ62924 standard; cDNA; 8920 BP.
XX
XX AC AAQ62924;
XX
XX 25-MAR-2003 (revised)
XX 06-DEC-1994 (first entry)
XX
XX Carbamoyl-phosphate-synthetase II.
XX
XX Carbamoyl-phosphate-synthetase II; CPSII; psCPSII gene; malaria; ss.
XX
XX Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX CDS 1226..8401
XX /*tag= a
XX
```

```

FT XX /EC_number= "6.3.5.5"
PN XX WO9412643-A1.
PD XX 09-JUN-1994.
XX 02-DEC-1993; 93WO-AU000617.
XX 03-DEC-1992; 92AU-00006206.
PR 16-DEC-1992; 92AU-00006380.
XX (UNIX ) UNISEARCH LTD.
PA Stewart TS, Flores MV, Osullivan WJ;
XX WPI; 1994-200271/24.
DR P-PSDB; AAR55694.
XX
XX Nucleic acid encoding carbamoyl phosphate synthetase II - isolated from
PT Plasmodium falciparum, used to develop prods. for the treatment of
PT malaria.
PT
PS Disclosure; Page 6-16; 31pp; English.
XX
XX The cDNA sequence encoding the carbamoyl-phosphate-transferase II (CPSII)
CC of Plasmodium falciparum was determined. The cDNA encodes a protein that
CC includes 2 insert sequences not found in other CPSII proteins. The first
CC separates the putative structural subdomain and the glutaminase subdomain
CC of the glutamine-amidotransferase subunit of CPSII, while the second
CC separates 2 ATP binding subdomains of the CPSII subunit, CP5a and CP5b.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 8920 BP; 3836 A; 774 C; 1232 G; 3078 T; 0 U; 0 Other;
SQ
Query Match 2.8%; Score 59.4; DB 2; Length 8920;
Best Local Similarity 45.7%; Pred. No. 0.052;
Matches 207; Conservative 0; Mismatches 246; Indels 0; Gaps 0;
QY 266 ATATATACGTGATGATGATGACACCATTTGTTGTCATTAACTTCCCACTCTATACATC 325
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 ATATATATATATATGTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 392
QY 326 AGTATCTCAAGTCGAATACATATCCATAGAGTGGTATTTGTAAGAAAAA 385
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 ATTAATATATTTCCCAATTAATGAAATACAAATTAATTTGATGTGACATTA 452
QY 386 AAAAGTGGTATACTGTTATATACATACACCGCTCTCGAATGCCCTCAACAATTTCTAGG 445
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 453 TATAGTTTTACACTTCTTATATATAAACCATCTCTATATATACACATATATATACT 512
QY 446 AGAAATGACGTGCTCTTTGTTTATTTATTTATTTATTTATTTATTTATTTATTTAA 505
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 513 CCCCAATATGTTGTTCTTATAATTTTATTTATATATTTATTTATTTATTTATTTATTTA 572
QY 506 ACACCTTCGATGCTCGCTTAAATTTTGAATGCGCTAAATTTCTTAATCATAAATCGTA 565
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 573 TTTATTTTTTTCTTAGTTTATAAATAGTAATCTTACTAATTTAAATAAAAAA 632
QY 566 AAGAAATTCGTCGAGCCACAGGGACATGCATAGGGCAGCTAGTTTACCTTTAAACCAT 625
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 633 AAAAAAAGAAAAAAGAAAAAATTTACATATGAAAAATGAACTTGTATATGTAAT 692
QY 626 CAAATATATTAATAGAAAGAACTTCCCTAAAGAACAAATTAATTAAGTGGATAA 685
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 693 TTATAAATATTTTAAACATATAAATGTATAAAAAAAGAAAAAATGGGAAAA 752
QY 686 AAAAGATAAGAGGTAGGACAGAAAAACGTAT 718
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 753 ATAAATAGATATATATATAAATATATATATAT 785
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 10
ABQ67005
```

```

ID ABQ67005 standard; DNA; 33053 BP.
XX
AC ABQ67005;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 35.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPTG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
WPI; 2002-500450/53.
XX
DR New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
PT
XX Claim 1; SEQ ID NO 35; 41pp + Sequence Listing; German.
XX
XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 33053 BP; 9206 A; 283 C; 6714 G; 16850 T; 0 U; 0 Other;
Query Match 2.8%; Score 59.4; DB 6; Length 33053;
Best Local Similarity 51.1%; Pred. No. 0.064;
Matches 166; Conservative 0; Mismatches 156; Indels 3; Gaps 1;
QY 1064 GTTAATATGATGTTGGCACTCTCGTATTAAATATGCCAGTGCACCTAGATAAAAA 1123
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 8962 GTGAAAAATTTAATAATAGTTTTTATAATATAAATAATTAGTTAAATAATTTAAAT 9021
QY 1124 ACATGATAGACATTTAGTTTAAACCTTGAAATGTTATTTGAACTCTTTTCGATTACGTGA 1183
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9022 ATATAATTTATATTTTGTAAATTAGAGAAATGTTTATTAATAATAAGATAATCTTAA 9081
QY 1184 TTGTTGATGGATTAATTTTGAAGATATTTATATATTTGAAGATGTTTATATATTAGA 1243
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9082 TTTGATTTTGAGTAAGTATTTTGTGAAATTTGGTTTGG---TGTTCGTTTTTTTTTA 9138
QY 1244 GTTTATATAGCAGAAAAATTTGATGTAGATGTGTCCTTTTCTAGTACTCTTTTGTGT 1303
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9139 TATTAATTGATAAAAAATTTTATTTATTGTAATTAATTTTGTTCGCGGTTTTTATC 9198
QY 1304 GCGTAGTCCTTCTCCTCATCTCTCTATCGAAGAAAAATCCAAATAGTTTAAAGAAATTTT 1363
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


CC	European Patent Office
XX	
SQ	Sequence 6045 BP; 2228 A; 31 C; 950 G; 2836 T; 0 U; 0 Other;
	Query Match 2.7%; Score 58.2; DB 6; Length 6045; Best Local Similarity 44.6%; Pred. No. 0.084; Matches 283; Conservative 0; Mismatches 343; Indels 9; Gaps 1;
Qy	856 TATTATATACTAAAGATCCCTCTGTGCTTTTGGCTTATTCCTGATATTAATCTAAC 915
Db	3012 TAAGATATATAAATTTTGATATTTTTTGGAATATATTTATTAATATATTTATA 3071
Qy	916 TTAAATAGTCTAAATAATATATATGCCTACCTATGTTCTACTGACCTCAGTCOCTAGT 975
Db	3072 TGAAAGGTTATTTAAAGTATTTTGGATATATAAAATTTTGTGATTTATATAGTAAT 3131
Qy	976 TAGCATATGCATATGTGAAAATGAGGCCCAAAATTTGAAGAGTTCCTCTCCTCGMA 1035
Db	3132 AAATATATATATGTATAGTATATAAATTTTAGAAATTTTATTTTGGATATTTTAAA 3191
Qy	1036 CTAACCTTATCTTACTGATGATGATTAATAATTCGAATGTTGGCACTCTCGTATTA 1095
Db	3192 TAAGGATTTAGTATATTTATATATATTTTTTTTGGTAGTTATTTTAAAAATA 3251
Qy	1096 AATATGCCAGTTGCACCCTAGATAAAAAACATGATAGACATTTAGTTTAAAACTTGAAT 1155
Db	3252 AAAATGTTAAT-----ATAAGGTTAGAGTTTTTATATGATATAATAGTAGATAATTT 3302
Qy	1156 GTTATTTGAACTCTTTGGATACGTGATGTTGTATGGATTAATTTTGAAGATATTA 1215
Db	3303 TATATATAAATGTTTGTAGATAGTTATAGATPACGATTTTTTTTTTAAAGAAGTTT 3362
Qy	1216 TATATTGAAGATGTTTATATATATATAGAGTTTATAPAGCAGAAAATPATGTAGATGT 1275
Db	3363 AAAAAATTAGATGATATATTTTATTTTTTTTGGABAAATATAGNGATATATTTTTTAT 3422
Qy	1276 TGTCTTTTGTAGTACTCTTTTTTGTGGTAGTCCCTTTCTCCTCACCTCCTATGAAG 1335
Db	3423 ATTTGTTTATAAATGAGTATTTATATAAATTATGAGTAATAATTAATGTTAATGATA 3482
Qy	1336 AAAAAATCAATAGTTTAGGAATTTTGTGTAATTCATAGTCTTTTTCGTAAACACAG 1395
Db	3483 AATATTTTTTTTATATAGGAATTTTTTATGATTTTTTAAATTTTTTATAGTATGTTTA 3542
Qy	1396 TTCATGTAGCTATCGTCATCATATTCCTCTTGGCAACAACAAAAAGATCGTTTIGTA 1455
Db	3543 GTTTAGTTTCTTAAAGGTAATGTGAAGTTTACGATATATTTAGGATCGAGTTTATAG 3602
Qy	1456 AAATTTAGTAGGCACTAAGTCGTCAATTTGTTGT 1490
Db	3603 GTTTTGTGAAGTTAAATTTATATAATTTATTTTTTTT 3637
RESULT 14	
ABZ10100	
ID	ABZ10100 standard; DNA; 8056 BP.
XX	
AC	ABZ10100;
XX	
DT	16-JAN-2003 (first entry)
XX	
DE	Haematopoietic cell proliferation disorder related DNA sequence #240.
XX	
KW	Human; haematopoietic cell proliferation disorder; cytostatic;
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW	cytosine methylation state; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200277272-A2.
XX	
PD	03-OCT-2002.
XX	

PF	26-MAR-2002; 2002WO-EP003401.
XX	
PR	26-MAR-2001; 2001US-0278333P.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI	Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI	Schwöpe I, Ziebarth H;
XX	
DR	WPI; 2003-018942/01.
XX	
PT	Detecting and differentiating between hematopoietic cell proliferative that disorders, comprises contacting a target nucleic acid with a reagent that distinguishes between methylated and non-methylated CpG dinucleotides.
PS	Claim 28; SEQ ID NO 240; 117pp; English.
XX	
CC	The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients
XX	
SQ	Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;
	Query Match 2.7%; Score 58.2; DB 7; Length 8056; Best Local Similarity 45.0%; Pred. No. 0.088; Matches 259; Conservative 0; Mismatches 313; Indels 3; Gaps 1;
Qy	830 AAACCATATCCAATTTCTCGCTATTTATATATAAATCAAGATCCCCTCTGTGCTTTG 889
Db	6875 AAAAAAAAAATGTTAATTTTTTTTTTTCGAATAAATCGAACGTTAAATTTTTTTTAAAT 6933
Qy	890 TCATTATTCGTGATATATAATCTAACTTAAATAGTCTTAAATAATATATATGTCCTACCTA 949
Db	6935 TTTTTTTTTTTTTTTTATTTAAAAATACGATTATATATAAAAAATATATCGATTTTAAAT 6999
Qy	950 TGTTTCTACTGCCTCAGTCCCTAGTACGTATATATGGACATATGTGAATAACGCCCAA 1009
Db	6995 TAAAAAATAAAAAAATAATTTATTTTCGTATTTTTTTTTTAAAAAATAAAAAATTTTANA 7054
Qy	1010 AATTTGAAGAGTTCCCTCTTCCTGCAACTAATCTTATCTTACTCAATTGAGCTATGTTAA 1069
Db	7055 TTAATTTAAATTTTAAATATTTTTTTTAAATATTAATTTTTTAAATTTATTAATTTTT 7114
Qy	1070 TATTGAATGTGGCACTCTCGTATTAATAATATGCCAGTTGCGACCTAGATAAAAAACATGA 1129
Db	7115 TTTTAAATATTAATAAAAAATATATATATAATTCGTAACGTAATTTATTTATTTTATTT 7174
Qy	1130 TAGACATTTTAGTTTAAACCTCGAATGTTATTTTGAATCTTTTGGATTACGTTGTTG 1189
Db	7175 TAGCTTTCGATTTTAAATTAATTTTAAATTTTATATATAATTTATTTATTTTAAAAATA 7234
Qy	1190 TATCGGATTAATTTTGAAGATATTTATATATTAATTAATTAATTAATTAATTAATTA--GAGCT 1246

Db 7235 TTTATATTTAAATATATAATATATATTTTTCGAAATTTTATTTATATATTTTACGA 7294
QY 1247 TATATAGCAGAAAATTCATGATGATGTTGTCCTTTTCTAGTACTCTTTTGTGCG 1306
Db 7295 TATATAAATAAATAATATTTTATTTTAAATTTTATTTAAATAATATTTAAATA 7354
QY 1307 TAGTCCTTTCTCTCATCTCTCTATGAAGAAAATCCAAATAGTTTAAAGAAATTTTGT 1366
Db 7355 AATTTTCTTTTATTTATTTTAAATAAATAATTCGAAAAAATAAATACGT 7414
QY 1367 GTAATTCATAGCTTTTTCGTACACACAGTTCAT 1401
Db 7415 ATATAATATACGTTTATTTAAATAAATAATTTAT 7449

RESULT 15
ABQ66998
ID ABQ66998 standard; DNA; 37515 BP.
XX AC ABQ66998;
XX DT 28-AUG-2002 (first entry)
XX DE Human angiogenesis associated polynucleotide SEQ ID NO 28.
XX KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.

XX OS Homo sapiens.
XX PN WO200246454-A2.
XX PD 13-JUN-2002.
XX PF 06-DEC-2001; 2001WO-EP014320.
XX PR 06-DEC-2000; 2000DE-01061338.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Schacht O;
XX DR WPI; 2002-500450/53.
XX PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.

XX Claim 1; SEQ ID NO 28; 41pp + Sequence Listing; German.
XX PS
XX CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 37515 BP; 11428 A; 425 C; 7165 G; 18497 T; 0 U; 0 Other;

Query Match 2.7%; Score 57.6; DB 6; Length 37515;
Best local Similarity 48.8%; Pred. No. 0.15;
Matches 156; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 1146 AACCTGAAATGTTATTTGAACTCTTTGGATTAGCTGATTTGTTGATGATTAATAATTTG 1205
Db 6823 AAGTATAAGTTTTTGTAGAGATATTTTGTGTTTTTTTGTAGTAAAGTTTATTAATTTTA 6882
QY 1206 AAGATATTTATATATATTGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAATATTG 1265
Db 6883 ATAAATATTAAAGGATGTTTATGTTTATTTGTAAGAAAGATAATTTATATTGATAAAGA 6942
QY 1266 ATGTAGATGTTGTCCTTTTGTAGTACTCTTTTGTGTCGAGTAGTCCTTCTCCTCATCC 1325
Db 6943 TTGTAGTATTTTGATATTTTGAATTAATTTTGGATTTTTTTGTTTATTTTAAATAAAA 7002
QY 1326 TCCTATGAAGAAAAATCCAAATAGTTTAAAGGAAATTTTGTGTAATTCATAGTCTTTTTC 1385
Db 7003 TGATTTGTGAGAAATGATATTTTGTAGTAAATATTTTATTTTAAATTGTTATTTAT 7062
QY 1386 GTAACACAGTTTCTATGTAGCTATCGTCAATCATATTCCTCTTTGCAACAAACAAAAGAT 1445
Db 7063 ATAAATTAATTAATTTGAATAATCGTATAAAGTTATTTTAAATAAATAAATAAAGT 7122
QY 1446 CGTTTTTGTAAAATTTAGTA 1465
Db 7123 AAATATTGTTGTAATTAATA 7142

Search completed: July 31, 2004, 04:28:03
Job time : 866 secs


```
Db 16175 GTATATATGTTTATAGTATTATATATATGTTTGTAAAGTATATATATATATATGTTAGTTATG 16116
Qy 1200 ATTTTGAAGATATTTATATATGAAGATGTTTATATATATATAGATTTATATATAGCAGAAA 1259
Db 16115 GGTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 16056
Qy 1260 ATATTGATGATGATG-TGTCTTTTGTAGTTTACCTTTTGTGTCGTCGTCCTTCTC 1318
Db 16055 GTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 15996
Qy 1319 CTCATCCTTCATGAGAAAATCAAAATAGTTTAAAGGAATTTTGTGTAATTCATAGT 1378
Db 15995 TATATATGATATATATGTCGATGAAATTAATTTTATTTTATTTTATTTTATTTTAAATTTAA 15936
Qy 1379 CTTTTTCGTAAACCAAGTCTTATGATGCTATGCTATCATATATTCCTTTTGCACAAACAA 1438
Db 15935 TTTTTTATTTTATTTTATTTTCAATTAATTTTAAATTTTATTTTATTTTATTTTATTTTAAATTT 15876
Qy 1439 AAAAGATCGTTTGTGTAATAATTTAGTAGGGCACTAAAGTCGTCATTTGTTGTCCTGTGCGA 1498
Db 15875 TTTTATTTATTTTATTTTATTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 15816
Qy 1499 AATCTAGCGTCTGTCATCCCAATATAGTTTGTGATTCGAGCTTCCCAAGATTAATATC 1558
Db 15815 AAATAAATTTTATTTTATTTATGTAATATTTTATTTTAAATTTT-----TTTAAAT 15762
Qy 1559 TTTTATAGTGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATATTTCT 1618
Db 15761 TTTTTTATTTTATGATATATATTTTATTTTAAATATATTTTCTTTTATTTTATTTTATTTT 15702
Qy 1619 AACATATACGCTTGTGTTTGGTAGGCTCTCGCTTTTGTAGACCAACCCCTTCTTAATG 1678
Db 15701 TTATGATATATATTTTATTTTATTTTAAATTTTATTTTCTTCTTTGTTTAT-TTT 15643
Qy 1679 TTTTGTGTCACCTTAGACATCCATATACGTTAGTGAGTGCAGTGTGCACCAAAATGG 1738
Db 15642 TTTTATATATATATTTTATTTTATTTTAAATTTTATTTTAAATTTTATTTTGTGATAATCTTT 15583
Qy 1739 TCCAAATATATTTTAAATTTTGGCCACAAACAAATTTTACAAACAAATTCACAAACAT 1798
Db 15582 TCATTTTATTTCTATCAAAATTTATATTTTATTTATTTTATTTTATTTTATTTTATTTTAAAT 15523
Qy 1799 GCATCGTTTCAATTTTATTTTCAATGGGTTTATTTGTTTCAATGTAATATTTCTGTTT 1858
Db 15522 TTTTCTCCTTTTATTTTATTTTATTTTAAATTTTATTTTAAATTTTATTTTATTTTATTTTCAAT 15463
Qy 1859 AACTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTGTATATAATTAACATTT 1918
Db 15462 TTTCTTTTCAATTTTAAATTTTGTTTTATATTTCTTTTATATTTTATTTTAAATATAACATATA 15403
Qy 1919 TATGGAA 1925
Db 15402 TATAAAA 15396
```

RESULT 4

```
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMIDIUM FALCIPARUM
; FILE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: JII-002CNCB
; CURRENT APPLICATION NUMBER: US/08/213.419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
```

```
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3
```

```
Query Match 2.4%; Score 51.2; DB 4; Length 6124;
Best Local Similarity 44.1%; Pred. No. 0.033;
Matches 357; Conservative 0; Mismatches 443; Indels 10; Gaps 3;
Qy 1114 AGATAAAAAACATGATAGACATTTAGTTTAAAACTTTGAAATGTTATTTGAACCTCTTTGG 1173
Db 1465 AAAAAATCAAGCTTATGTTCTGATATTTCAATGGATATATAAAATTAATTTTITTA 1524
Qy 1174 ATTACGTGGATTGTTGATGCAATTAAT-----TTTGAAGATATTTATATATTTGAAGAT 1227
Db 1525 ACTCTAGAAATATTTAATGAATTAATACGACCTATATAATCTTGTATGTAAACT 1584
Qy 1228 GTTTATATATATTAGAGTTTATATAGCAGAAAAATATTGATGTAGATGTTGCTTTTGTGA 1287
Db 1585 TTTTATATTGGAAGAAAAAATAATTTATATATATTTTATTTTAAATTTTATATATA 1644
Qy 1288 GTTACTCTTTTGTGCGTAGTCCTTTCTCCTCATCTCCTATGAAGAAAAATCCAAAT 1347
Db 1645 TATTATAATTACAATTTTTTTTTTTTTTTTTTTTTTTTAAACCGAATGAATATATTT 1704
Qy 1348 AGTTTAAGAAATTTTGTGTAATTCATAGTCTTTTTCGTAACACAGTCTCTATGTAGCT 1407
Db 1705 TATGAGATCAAAATTTTAAATTTTATTTTTCGGTTTAAATATTTTTTTTGTGTCGA 1764
Qy 1408 ATCGTCATCATATTCCTCTTTGCAACAAACAAAAAGATCGTTTTGTAAAAATTTAGTAGG 1467
Db 1765 CGATAACTAATCCCTTGTATTTTAAAAAAAATGTACACATGTACATATGATATAT 1824
Qy 1468 GCATAAAGTCGTCATTTGTTGTCCTCTCGAAATCTAGCGTCTGTGCATCCACAAATAAG 1527
Db 1825 ATATATATATATATGATTTCTTACAAATTAATATATTTTAGAATATTTAAATTTCT 1884
Qy 1528 TTGTTGATTCGAGCTTCCGAGATTAATCTTTTGTAGATGGTCAATGAAGATTTCTAA 1587
Db 1885 TAGAATAACTATATTTTGAATAATCCGAAATCTTAAATGTTTACAAAAATGAGAAATATAA 1944
Qy 1588 CTTGCTATAGAGTGTATCCATATAATTTCTAATATACGTCT--TGTTTTTGGTAGG 1644
Db 1945 AACTCACTTATATATATATATATATATATATATATATATATATATTTATATTTTAATGCA 2004
Qy 1645 CTCTGCGTCTTTTGAAGACCCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1704
Db 2005 CTTTAAATATATGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 2064
Qy 1705 ATACGTTACGTGAGTCGAGTTCGACCAAAATGGTCCAAATATTAATTTAAATTTTGGCCAC 1764
Db 2065 ATATTTTTTTTGAATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 2124
Qy 1765 AAAACAACATTTTACAAACAAATTTCAACAAACATGATCGTTTCAAAATTTTATTTATTTCA 1824
Db 2125 AGAATGTACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2184
Qy 1825 ATGCGGTTATTTGTTTCAATTTGTAATAT-TCTGTTTAACTCACTGACGAATTTTAAATTT 1883
Db 2185 ATAAATATATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2244
Qy 1884 TTCAAGAGAGAACATTTTGTGATATAATAA 1913
Db 2245 AAAATTTAAATTTAGATTTGTCCAAAAAATAA 2274
```



```

; PRIOR APPLICATION NUMBER: DE 1003529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 33
; LENGTH: 5455
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated
US-10-204-708-33

```

Query Match	2.3%;	Score 50;	DB 4;	Length 5455;
Best Local Similarity	46.4%;	Pred. 0.059;		
Matches 273;	Conservative 0;	Mismatches 305;	Indels 10;	Gaps 3;
QY	1041	TCTTATCTTACTCATTCGACTATGTTAAATATCGAATGTGGCACTCTCGTATTAAATAT	1100	
Db	536	TTTTAATCGAGTTTTTTTTTTCGTTTGATTTTTTTTATATTATTTTTTGTGGTAGTTAC	595	
QY	1101	GCCAGTTGCACCTAGATAAAAAACATGATAGACATTAGTTTTAAAACTTTGAAATGTTAT	1160	
Db	596	GTTAAATTTACGTTTTTTTTCGTTATTTTTTTTTTTTGTTTTTTTTTTTAGTTTTCGGGAATAT	655	
QY	1161	TTGNACTCTTTGGATTACGTGGATTGTGTATGGATTAAATTTTGAAGATATTATATAT	1220	
Db	656	TTAATGTTATTCGAATTAATTTTTTAGTTTTTTGTTTTTTAAATTTTCGATTTAGTTTTTGT	715	
QY	1221	TGAAGATGTTTATATATATTAGAGTTTATATAGCAGAAAAATATTGATGTAGATGTTGTCC	1280	
Db	716	T----TGTGTTTCGTTTTTTTAAAGTTTTTAAATTTATTGTTTTTATTATTAGTT---TTT	770	
QY	1281	TTTTGTAGTTACTCTTTTTTGTGGGAGTCCTTTTCCTCATCCCTCTATGAAGAAAA	1340	
Db	771	TTTTAATATATCGTTTTTTTTTGAACGCGTTTTTATTATATATATATATATATATATA	830	
QY	1341	TCCAAATAGTTTTAAGGAAATTTTTTGTGTAATTCATAGTCTTTTTTCGTAACCAAGTCTA	1400	
Db	831	TATATATATATATATAATTTTTTTTATGTATATTATTATTTTTTAGATTTGTAGGGATATA	890	
QY	1401	TGPAGCTATCGTCATCATATTCCTCTTTGCAACAAACAAAAAGATCGTTTTCTAAAAAT	1460	
Db	891	TGTTTTAAATGAGGTTTTTTTTTTTTTTTTTAAATTTTGTTCGGTTTTTATTGTTTTTTT	950	
QY	1461	TAGTAGGGCACT-----AAAGTCGTCATTGTTGTCCTGTCGAAATCTAGCGTTCTGTCA	1515	
Db	951	TATTAGTTCGTTTTTAAATTTTATTTTAGTFTTTATTTTTTGGGAATTTTTTTTTTTTTTA	1010	
QY	1516	TCCACAAATAAGTTGTTTGATTTCGAGCTTCCAAAGATTAAATCTTTTTTTTAGATGGGTCAT	1575	
Db	1011	TTTTTTTTTTTTTTAGTTTTGTTAGAAATTCGATTTTTTATATAATTTTTTAGGTTGATTTT	1070	
QY	1576	GAAGATTTCTAAGTTCGTATACGAGTGTATCCATATAAATTTCTAACAT	1623	
Db	1071	AGAAATTTTATATTTGTTAATAAGTAAATGTTTTTTTTTTTTTTTTTAAAAAT	1118	

```

RESULT 8
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07

```

```

; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

Query Match	2.3%;	Score 49.2;	DB 4;	Length 640681;
Best Local Similarity	44.2%;	Pred. No. 0.33;		
Matches 252;	Conservative 0;	Mismatches 313;	Indels 5;	Gaps 1;
QY	127	TGAATCTTTTCATTTCTTGAATGGATATTATACATTTTTATCAAAAAAAGTCAAGTTCT	186	
Db	32547	TAAATTTATAAAGTTTTTATTGAAATGTTTTGAAATATATTCAAATCGAAAATAAAAAATTTTT	325488	
QY	187	ACCAAAGCACAGGAGTTTAAACAACCTTGCTGTGTCAAATGCTAATTTTAAAGCCCTTAATCTTAT	246	
Db	325487	ATTTTATCGATTTCGATGCAAAATCTTTTTTATATAAAATAGAAATAAAAAAAGAAT	325428	
QY	247	GATTTCCCTTTTCTTCAGGATATATACGTATATTCGATGATGCACCCATTTGTTTGCATT	306	
Db	325427	TACATCCTTATTTATTTATATATGATAATTTTTTTGTTCTTTA-----TTTTACTATTTTT	325373	
QY	307	ACTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAGTGGTA	366	
Db	325372	ATTATATACTATAATTATCATTTAGTATATTTAAAAAATATAAAGAAAAATATCAGTA	325313	
QY	367	TATTTGTGAAAAAATAAAAAAAGTGGTATATCTGTATATACAATACCACGCTCTCGAAT	426	
Db	325312	TTTTTATTTATATAAAATGTATTTTATTAAAAATATAAAATATTTATAGTATAAAATTTA	325253	
QY	427	TGCTCAACATTTCTAGGAGAAATCGACGTGTCTCTTGGTTTTATTTTATCTTTAAT	486	
Db	325252	ATATAATAAATGTAAAAATAAAATAATTTATTTATATTTTTTTTAAATACACATTTTAAA	325193	
QY	487	AACATACCTCTATTTTTTAAACACCTTCGATGTCTCGTTTAAATTCGAATGTCCTTAAT	546	
Db	325192	ATATTTATTTTAAAGAAATATTTTTTAAATTTATAAAAAATATAATGATGAATAAAATTTAGTT	325133	
QY	547	TCTCTAATCATAAATCGTAAAGAAAAATTCGTGGAAGCCACAGGACATGCATAGGCACG	606	
Db	325132	TAAATTTAAATTAATAAAATTTATTACTTTTATTTTAAAGAAATACGTTCTATGAAAAATACAAA	325073	
QY	607	TAGTTACCTTTTAAACCATCAAAAAATATTTAATAGAAAGGAACCTTCCTTAAAGACA	666	
Db	325072	AAATAAAATTTTAAAAAACHAAAAAATTTATTTATTTTAAATATCAAAAAATTACTAGATCT	325013	
QY	667	ATTTAATAAAGTGGATAAAAAAGATAAGA	696	
Db	325012	ATTTTATAATGTAAATAATCAAAAAAATA	324983	

RESULT 9
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731

```

: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
: TITLE OF INVENTION: by Assessing DNA Methylation
: FILE REFERENCE: 5013.1012
: CURRENT APPLICATION NUMBER: US/10/204,708
: CURRENT FILING DATE: 2003-05-06
: PRIOR APPLICATION NUMBER: PCT/EP01/03971
: PRIOR FILING DATE: 2001-04-06
: PRIOR APPLICATION NUMBER: DE 10019058.8
: PRIOR FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: DE 10019173.8
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: DE 10032529.7

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

Query Match 2.3%; Score 48.6; DB 4; Length 11049;
Best Local Similarity 44.2%; Pred. No. 0.15;
Matches 247; Conservative 0; Mismatches 309; Indels 3; Gaps 1;

QY 1086 TCTCGTATTAAATATGCGAGTTCACCTAGATATAAAAAAATCATAGACATTTAGTTTAA 1145
DB 8407 TTTTATTTTAAAGAAATTTTATGCTGTATATAAATAGAAAAATGAATAGATTG 8466

QY 1146 AACTTGAAATGTTATTGAACTCTTTGGATTAGCTGGATTGTTGTATGATTAATTTG 1205
DB 8467 TTAATTTAGTTAGTATAGATTTTGGATTGGAATTTTGTGTTTGTGTTGAAAGTG 8526

QY 1206 AAGATATTATATTTGAAGATGTTTATATATATATAGATTATATAGCAGAAAAATTG 1265
DB 8527 AGGTGAGTTGAGTTTATTTTGTGAAATAATATAGTAGTTTAAAGAAAGAAATTTT 8586

QY 1266 ATGTAGATGTTGCTTTT---TCGTAGTTACTCTTTTGTGCTAGTCTCTCTCTCA 1322
DB 8587 ATGTGTTGTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATG 8646

QY 1323 TCCTCCTATGAAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATGCTTT 1382
DB 8647 TTTAGGTATTTGTTAGAGTTTGGGAATTTTGGAAATTTGTTTATTTTGTGATGTTT 8706

QY 1383 TTGCTAACACAGTTCTATGATAGTATCGTCATCATATTCCTCTTTCACACACAAAA 1442
DB 8707 ATGTTATTTAGATTTTATTTTGGATTTAGATTTGTTATTTTGTGATGTTGTAATGG 8766

QY 1443 GATCGTTTGTAAATTTTAGTAGGCGCACTAAAGTCTGCTATTGTTGCTCTGCGAAATC 1502
DB 8767 TATGATATATAGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGAGT 8826

QY 1503 TAGCGTTCTGTCATCCAAAAATAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1562
DB 8827 TTTATTTTGTGATTGAATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTAT 8886

QY 1563 TTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAACA 1622
DB 8887 TTAGATTTAGTTTGAATTTATTTTATTTTATTTTATTTTATTTTATTTATTTATTA 8946

QY 1623 TATACGCTCTGTTTGTGTTTGGT 1641
DB 8947 GATTATGTTGTTTGTGTTTGGT 8965

RESULT 10
US-09-620-312D-481
; Sequence 481, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrei
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 481
; LENGTH: 6765
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(3105)
US-09-620-312D-481

Query Match 2.3%; Score 48.4; DB 4; Length 6765;
Best Local Similarity 49.2%; Pred. No. 0.14;
Matches 127; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 994 TGAATATGAGCCCAAAATTTGAAGAGTTCTCTCTCTGCACTAACTCTTATCTTACTC 1053
DB 5350 TTAATCTTACCTAAAGAGTCTGAAATGCTTTTGTAGTACTTTTTCATATTTATACAT 5409

QY 1054 ATGAGCTATGTTAAATATATGATGTTGGCACTCTCGTATTAATATGCCAGTTGCACCT 1113
DB 5410 TATTTCCGAATGTATCATGTAATAATTTTATTTAGTTATATAAGATATCTTATTCATT 5469

QY 1114 AGATAAAAAACATGATAGACATTTAGTTTAAAACTTGAAATGTTTATTTGAACCTTTGG 1173
DB 5470 TAAATAAAATTAACATATAAAGTGAATGCTGAAATGCTGAAATGCTGAAATGCTGAA 5529

QY 1174 ATTACGTTGATGTTGTTGATGATTAATAATTTTGAAGATATTTATATATTTGAAGATGTTAT 1233
DB 5530 TTAACACTGACTTAATGATGCTGCTCAATTTTGTGCTCTTATTTTCTGATCATTC 5589

QY 1234 ATATATTAGATTTATAT 1251
DB 5590 TCCTTTTATAGTTTACAT 5607

RESULT 11
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1


```

; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 80
; LENGTH: 8961
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (5096)
; OTHER INFORMATION: n is a or g or c or t
US-10-204-708-80

Query Match      2.2%; Score 47.4; DB 4; Length 8961;
Best Local Similarity 46.3%; Pred. No. 0.26;
Matches 156; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 972 TAGTTAGCTATATGGACATATGTAAGAAATGACGCCCAAAATTTGAAGAGTTCCTCTCTCT 1031
DB 5793 TCGTTTATAGATGGATATTAATTTTTTTTTTTTTTATATTGGAAGTTATTTTATAAT 5852

QY 1032 GCACTAATCTTATCTTACTCATCTGAGCTAAGCTTAAGTAAATATGCAATGTTGGCACTCTCGT 1091
DB 5853 AATTTTATTTTATTTATTTATTTGATATTTTATTAATATTTTATTAATTTTATTTATTT 5912

QY 1092 ATTAAATATGCCAGTTGCACCTAGATAAAAAACATGATAGACATTTAGTTTAAACCTTG 1151
DB 5913 ATTAAATATATTATATATTTTAAATATATTTAGTGTATATTTTATTAATGTTTATTT 5972

QY 1152 AAATGTTATTTGAACCTTTTGGATACGTGGATTTGTTGATGAATTAATTTTGAAGATA 1211
DB 5973 TTAATTTATTTTATTTTATTTTTCGAGTTAGGTATTTTATTAATATTTATTTTATTT 6032

QY 1212 TTTATATATTGAAGATCTTATATATATAGTTATAGTTTATATAGCAGAAAATATTGAITAG 1271
DB 6033 TTTTATTTTATAGCGTGTAGTTTATTTTAAATTTAGATAGGAGGGTGTAGGTAGG 6092

QY 1272 ATGTTGTCCTTTTCTAGTTACTCTTTTCTTGTTCGTA 1308
DB 6093 TTAATTTAGCTTTTCTGTTTCTGTTTATTTGTA 6129

RESULT 14
US-10-204-708-60
; Sequence 60, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 60
; LENGTH: 6156
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;

US-10-204-708-60
Query Match      2.2%; Score 47.2; DB 4; Length 6156;
Best Local Similarity 44.1%; Pred. No. 0.26;
Matches 240; Conservative 0; Mismatches 303; Indels 1; Gaps 1;

QY 765 GAGACGCTGACTGATTTTCTTTCTTTCTTAAAGAACGCTGTTGCTGCTTACAG 824
DB 4302 GAGGGGGTGGGATTTTTTTTTTAAAAATAATTTAGAAATAAAATTTTAGTTTGA 4361

QY 825 GGTCAAAACCATATCCCAATTTGTCCTTATTTATATACTAAAGATCCCTCTTGTG 884
DB 4362 AAAATATATTGGAAGATTGGTTATAAAATATTTTGAAGAGAAATTTTAGGTATTTT 4421

QY 885 CTTTCTCTTTTATTCGTGATATATAATCTAACTTAAATAGTTCTTAAATATATATGCTCT 944
DB 4422 AGGTGTTTTTTTTTGAATTTTATATAAGATTTATTTATTTTATTTGTTATTAATAAAT 4481

QY 945 ACCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGGACATATGTGAAAATGACG 1004
DB 4482 TAGGAAGTGTATAGTAATTTTTTAAATATTTAAAGTAATTTTTTAAATAAGTGGATATAGT 4541

QY 1005 CCCAAATTTGAAGAGTTCCTCTCTGCAACTAACTCTTATCTTACTCTCATTCGAGCTATG 1064
DB 4542 GTTATATGTTATTAATTTTATTTTATTTATTTATTTGTAATTTGAAATTTATTTATTTTATA 4601

QY 1065 TTAATATTTGAATTTGGCACTCTCGTATTTAAATATGCCAGTTGCCACCTAGATAAAAA 1124
DB 4602 TTTTCTTTGAAATGAAATTTATTTGTTTTTATTTTGGTTTTTATTTTATTTTATATAATAA 4661

QY 1125 CATGATAGACATTTAGTTTAAACCTTGAATGTTATTTGAACTCTT-TGGAATACGTGA 1183
DB 4662 GTTTATTTAAATATA-TTAAAGAGTTAGAAATTTGATTTTAAATTTTAAATATTTATTTTATA 4721

QY 1184 TTGTTGTAUGATTAATTTTGAAGATTTTATATATTTGAAGATGTTTATATATATATAGA 1243
DB 4722 TTGATGAGAGTTCGTTGATGGATTTTAAATTTAGAAATGTTATTTTATATATATATAA 4781

QY 1244 GTTTATATACAGAAAATA-TGATGTAGATGTTGCTCTTTTGTAGTTACTCTTTTGTGTT 1303
DB 4782 AGTAATTTAAATTTATATAGCGATTTTTTTTTTTTAGAGTTTGTGTTTTTAAATATTTT 4841

QY 1304 GCGT 1307
DB 4842 ACGT 4845

RESULT 15
US-10-204-708-6
; Sequence 6, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 6
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-6

Query Match      2.2%; Score 46.8; DB 4; Length 6669;
Best Local Similarity 43.9%; Pred. No. 0.32;
Matches 339; Conservative 0; Mismatches 427; Indels 6; Gaps 3;

QY 1117 TAAAAAACAATGATAGACATTTAGTTTAAACCTTGAATGTTTATTTGAACTCTTTGGATT 1176
Db      |||||
QY 3438 TTAATAATCGTTTTTTAGTTTGTAGTATGCTTAATATTTGGGTTTATTAATTTTATATG 3497
Db      |||||
QY 1177 ACGTGGATGTTGTATGATTAATAATTTGAAGATATTTATATATTGAAGATGTTTATATA 1236
Db      |||||
QY 3498 TGGGGAGTTAATTTTGTGTTTAAATATCTAGGATGTTTAGTAGTATTTTGGTTTATTT 3557
Db      |||||
QY 1237 TATTAGAGTTTATATAGCAGAAATATTGATGCTAGATGTTGCTCTTTTGTAGTTACTCTT 1296
Db      |||||
QY 3558 TATTAGA--TTATAGTAGGATATTCGTTTTTATTTTATTTTATTTTATTTAGTTGTGATTAGT 3615
Db      |||||
QY 1297 TTTTGTGCGTAGTCTTTTCTCTCATCTCTATGAAGAAATCCCAATAGTTTAAAG 1356
Db      |||||
QY 3616 GAAATATGTTTGTAGATATTTGTAGTATGTTTTTGGAGGTAGATATTTTGTAGTTAAG 3675
Db      |||||
QY 1357 AAATTTTGTGTAATTCATAGTCTTTTTCGTAACACAGTTCTATGTAGTATCGTCATC 1416
Db      |||||
QY 3676 AAAGTTTTAGAAAGTAGTATTTATTTTAAATGAAATATTTATATCGATGTGTGAGTA 3735
Db      |||||
QY 1417 ATATTCCTCTTTGCAACAAACAAAGATCGTTTTTGTAAATTTAGTAGGCACTAAAG 1476
Db      |||||
QY 3736 TTTTTTTTTATAAATTTAGTATTTTGTGTTGATTTTTTGTGTTTTTATTTATGATGATAAAA 3795
Db      |||||
QY 1477 TCGTCATTTGTGTCCTGCGAATCTAGCTTCTGTCATCCACAAATAAGTTCTGTTGAT 1536
Db      |||||
QY 3796 ACGTTAGTTATGTTTTTTTATTTTAAATTTTGTATTTATTTATTTATTTTATTTATGTT 3855
Db      |||||
QY 1537 TCGAGCTTCCAAGATTAATAATCTTTTTPAGATGGGTATGAAGATTTCTAACTTCGTATA 1596
Db      |||||
QY 3856 TATTAATTTCTGTAATGTTGTTGTTATTTATTTTGTGATTTTGTATGTTTTTTA 3915
Db      |||||
QY 1597 CGAGTGATCCATATAATTTCTACATATACGTTCTGTTTTTGTAGGCTCTGCGTCTTT 1656
Db      |||||
QY 3916 GATAGTTGTTTTCGTAATTTATTTTATTTA--TTGTTGTTATAAATTTTTTGAAGTTTT 3974
Db      |||||
QY 1657 TGAGACCAACCCCTTGCTAATGTTTTTGTGACCTTAGACAATCCATAATACGTTACGTG 1716
Db      |||||
QY 3975 ATATTATATTTTGTAGTTTGAAGTTGAGATAATTTTATTTTATTAATAATTTTAGATG 4034
Db      |||||
QY 1717 AGTCGRAGT---TGCACCAAAATGGTCCAAATATAATTTAAATTTGGCCCAAAACAACA 1773
Db      |||||
QY 4035 GGAGGAGATATAATATAAATAAAGTTTTTTTATGTTTAAATGTTTTTATTTTGTGCTTT 4094
Db      |||||
QY 1774 TTTTACAAACAAATTCACAAACATGTCATCGTTTCAAAATTTTATTTTCAATGGCGTTA 1833
Db      |||||
QY 4095 ATTATTTTATTTTAAATAGATATATATCGTGTATTTTTTTTTTATGAGGTTTATTA 4154
Db      |||||
QY 1834 TTGTTCAATGTAATATCTGTTTAACTCACTGACGAATTTTTTAAATTTTT 1885
Db      |||||
QY 4155 TTTATTTATGATTTTTTTTTTATAGTATTTTTTTCGTTATTTTGTAGTTTTTT 4206
Db      |||||

```

Search completed: July 31, 2004, 08:23:31
Job time : 171 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 06:47:54 ; Search time 975 Seconds
(without alignments)
10807.024 Million cell updates/sec

Title: US-10-027-880-1

Perfect score: 2149

Sequence: 1 gtcaagtgtgatatattgt.....actaatgcttaataatcag 2149

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 145 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2149	100.0	2149	15	US-10-027-880-1
2	2149	100.0	4280	15	US-10-027-880-2
3	2149	100.0	4309	15	US-10-027-880-4
4	2149	100.0	4413	15	US-10-027-880-3
5	63.6	3.0	7008	13	US-10-221-714A-254
6	60.6	2.8	396	17	US-10-021-323-17801
7	60.2	2.8	6175	15	US-10-311-455-1280
8	59.8	2.8	7597	15	US-10-311-455-986
9	59.6	2.8	5822	15	US-10-311-455-1070
10	59.4	2.8	33053	17	US-10-433-793-35
11	58.6	2.7	5413	13	US-10-221-714A-418
12	57.6	2.7	37515	17	US-10-433-793-28
13	56.8	2.6	8277	15	US-10-311-455-1301
14	56.6	2.6	6292	13	US-10-221-714A-461

15	56.6	2.6	3673778	15	US-10-312-841-1	Sequence 1, Appli
16	56.6	2.6	3673778	15	US-10-312-841-2	Sequence 2, Appli
17	56.4	2.6	15951	15	US-10-311-455-1653	Sequence 1653, Ap
18	56.4	2.6	15951	15	US-10-240-485-133	Sequence 133, App
c 19	56.2	2.6	17280	13	US-10-221-714A-498	Sequence 498, App
c 20	56	2.6	495	17	US-10-021-323-2679	Sequence 2679, Ap
21	56	2.6	37973	15	US-10-311-455-2170	Sequence 2170, Ap
22	55.8	2.6	6076	13	US-10-221-714A-386	Sequence 386, App
c 23	55.8	2.6	3673778	15	US-10-312-841-2	Sequence 2, Appli
c 24	55.6	2.6	393	9	US-09-960-352-4582	Sequence 4582, Ap
25	55.6	2.6	7072	13	US-10-221-613-348	Sequence 348, App
26	55.4	2.6	12237	15	US-10-311-455-2331	Sequence 2331, Ap
27	55.4	2.6	15224	13	US-10-221-613-223	Sequence 223, App
28	55	2.6	11745	15	US-10-240-453-206	Sequence 206, App
c 29	55	2.6	113515	15	US-10-311-455-2147	Sequence 2147, Ap
30	54.8	2.6	5297	13	US-10-221-714A-235	Sequence 235, App
31	54.8	2.6	14919	13	US-10-221-714A-228	Sequence 228, App
32	54.6	2.5	8244	13	US-10-221-714A-117	Sequence 117, App
33	54.6	2.5	8244	15	US-10-311-455-965	Sequence 965, App
34	54.6	2.5	8244	17	US-10-433-793-61	Sequence 61, Appl
35	54.6	2.5	13131	17	US-10-240-589C-57	Sequence 57, Appl
36	54.4	2.5	5273	15	US-10-311-455-848	Sequence 848, App
37	54.2	2.5	4661	17	US-10-433-793-51	Sequence 51, Appl
38	54.2	2.5	6191	15	US-10-311-455-1190	Sequence 1190, Ap
39	54	2.5	6223	13	US-10-221-613-135	Sequence 135, App
40	54	2.5	37515	17	US-10-433-793-27	Sequence 27, Appl
41	53.8	2.5	7498	15	US-10-311-455-230	Sequence 230, App
42	53.8	2.5	7851	15	US-10-311-455-1733	Sequence 1733, Ap
43	53.6	2.5	6665	15	US-10-239-676-4	Sequence 4, Appli
44	53.6	2.5	6665	15	US-10-311-455-56	Sequence 56, Appl
45	53.6	2.5	6665	15	US-10-240-453-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-027-880-1
; Sequence 1, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2149
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-027-880-1

Query Match 100.0%; Score 2149; DB 15; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTGCAATTGTGATATATTGTAAGCAATCTGAAGAAGATAAGTGGGATATAAACAACCG	60
Db	1	GTGCAATTGTGATATATTGTAAGCAATCTGAAGAAGATAAGTGGGATATAAACAACCG	60
Qy	61	GCGAAGTACAAAGTTCTACCTTTTTTGGCATGGAACCACTTTTGGAGTTACTTTGT	120
Db	61	GCGAAGTACAAAGTTCTACCTTTTTTGGCATGGAACCACTTTTGGAGTTACTTTGT	120
Qy	121	AATTCCTGAATCTTCAATTTCTTGAATGTATTTATCAATTTTATCAAAAAAAGTACA	180
Db	121	AATTCCTGAATCTTCAATTTCTTGAATGTATTTATCAATTTTATCAAAAAAAGTACA	180

; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2

; LENGTH: 4280
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Construction
; OTHER INFORMATION: promoter + sequence coding for the gus gene
US-10-027-880-2

Query Match 100.0%; Score 2149; DB 15; Length 4280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAATGTCGATATATTGTAAGCAATCTGAAAGAAATGAAGTGGGATATATAACACCG 60
DB 1 GTCGAATGTCGATATATTGTAAGCAATCTGAAAGAAATGAAGTGGGATATATAACACCG 60

QY 61 GCGAAGTACAAAGTCTACCTTTTGGCATGGAACCAATGTTTTAGGATTTACTTTGT 120
DB 61 GCGAAGTACAAAGTCTACCTTTTGGCATGGAACCAATGTTTTAGGATTTACTTTGT 120

QY 121 AATTCCTGAATCTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACA 180
DB 121 AATTCCTGAATCTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGTACA 180

QY 181 AGTTCCTACCAAGCACAGAGTTAAACAACTTGTGTCAAAATGCTAATTTAAAGCCTAA 240
DB 181 AGTTCCTACCAAGCACAGAGTTAAACAACTTGTGTCAAAATGCTAATTTAAAGCCTAA 240

QY 241 TCTTAGAATTCCTTTTCTCACGATATATCTGATATGATATGCAACCCATTTGTTTG 300
DB 241 TCTTAGAATTCCTTTTCTCACGATATATCTGATATGATATGCAACCCATTTGTTTG 300

QY 301 TCATTAACCTCCACTCTATACATCAGTATCTCAAAAGTCGAATAACAATATCCATAAGAA 360
DB 301 TCATTAACCTCCACTCTATACATCAGTATCTCAAAAGTCGAATAACAATATCCATAAGAA 360

QY 361 GTGGTATATTGTGAAAAAAGAAAAAGTGTGATCTGATATGATATCAATACCAACGTC 420
DB 361 GTGGTATATTGTGAAAAAAGAAAAAGTGTGATCTGATATGATATCAATACCAACGTC 420

QY 421 TCGAATTCGCCCTCAACAATTTCTAGAGAAATGGAAGTGTCTCTTTGTTTATTTTAT 480
DB 421 TCGAATTCGCCCTCAACAATTTCTAGAGAAATGGAAGTGTCTCTTTGTTTATTTTAT 480

QY 481 CTTAATACACTCTATATTTTAAACACTTCGATGTCGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 CTTAATACACTCTATATTTTAAACACTTCGATGTCGATGTCGCTGCTGCTGCTGCTGCT 540

QY 541 TAAATTTCTCTAATCATATAATCGTAAGAAATTTCTGCGAAGCCACAGGACATGCGATAG 600
DB 541 TAAATTTCTCTAATCATATAATCGTAAGAAATTTCTGCGAAGCCACAGGACATGCGATAG 600

QY 601 GGCACGTAGTTACCTTTAAACACCATCAAAAAATATTAATAGAAAAGAAACTTCTTAAA 660
DB 601 GGCACGTAGTTACCTTTAAACACCATCAAAAAATATTAATAGAAAAGAAACTTCTTAAA 660

QY 661 AGAACAAATTAATAAGTGGATTAAGAAATGATAGAGGTAGGAGAGAAACGTTATGG 720
DB 661 AGAACAAATTAATAAGTGGATTAAGAAATGATAGAGGTAGGAGAGAAACGTTATGG 720

QY 721 CCGGACTCGTAACAGGACGTCGCCACCTCGCGAGCGGAGACGCTGATGATT 780
DB 721 CCGGACTCGTAACAGGACGTCGCCACCTCGCGAGCGGAGACGCTGATGATT 780

QY 781 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTGCTTACAAGGGTCAAAACCATATCC 840
DB 781 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTGCTTACAAGGGTCAAAACCATATCC 840

QY 841 AATTTGTTCTGCCCTATTATTATATAAATGATCCCTCTCTGCTGCTGCTGCTGCTGCT 900
DB 841 AATTTGTTCTGCCCTATTATTATATAAATGATCCCTCTCTGCTGCTGCTGCTGCTGCT 900

QY 901 GATATATAATCTAACTAAATAGTTCTAAATATATATGCTCTACCTATGTTCTACTG 960
DB 901 GATATATAATCTAACTAAATAGTTCTAAATATATATGCTCTACCTATGTTCTACTG 960

QY 961 ACTCAGTCCCTAGTTAGCTATATGACATATGTAAGAAATGACGCCCAAAATTTGAAG 1020
DB 961 ACTCAGTCCCTAGTTAGCTATATGACATATGTAAGAAATGACGCCCAAAATTTGAAG 1020

QY 1021 TTCTCTCTCTGCACTAACTCTTATCTTACTCATTTAGCTATGTTTAAATATTGAATGTT 1080
DB 1021 TTCTCTCTCTGCACTAACTCTTATCTTACTCATTTAGCTATGTTTAAATATTGAATGTT 1080

QY 1081 GGCACCTCTGATATTAATAATGCAAGTTCACCTAGATAAAAAAACAATAGATAGACATTTAG 1140
DB 1081 GGCACCTCTGATATTAATAATGCAAGTTCACCTAGATAAAAAAACAATAGATAGACATTTAG 1140

QY 1141 TTTAAAACTTGAAATGTTTAACTCTTTGCAATACGTTGATTTGTTGATGGATTTAA 1200
DB 1141 TTTAAAACTTGAAATGTTTAACTCTTTGCAATACGTTGATTTGTTGATGGATTTAA 1200

QY 1201 TTTTGAAGATATTTATATATTTGAAGATGTTTATATATATAGATTTTATATAGCAGAAA 1260
DB 1201 TTTTGAAGATATTTATATATTTGAAGATGTTTATATATATAGATTTTATATAGCAGAAA 1260

QY 1261 TATTGATGATAGTTGTCCTTTTGTAGTACTCTTTTGTGCGTAGTCTCTTCTCTCT 1320
DB 1261 TATTGATGATAGTTGTCCTTTTGTAGTACTCTTTTGTGCGTAGTCTCTTCTCTCT 1320

QY 1321 CATCTCTCTATCAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1380
DB 1321 CATCTCTCTATCAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1380

QY 1381 TTTTCGTAACCAACAGTCTATGATAGTATCGTCATCATATTCCTCTTTTGCACACAAAA 1440
DB 1381 TTTTCGTAACCAACAGTCTATGATAGTATCGTCATCATATTCCTCTTTTGCACACAAAA 1440

QY 1441 AAGATCGTTTGTGAAAAATTTAGTAGGCACTAAAGTGTCTATTTGTTGCTGCGAAA 1500
DB 1441 AAGATCGTTTGTGAAAAATTTAGTAGGCACTAAAGTGTCTATTTGTTGCTGCGAAA 1500

QY 1501 TCTAGCGTTCGTCTCATCCACAAATAAGTTGTTGATTCGAGCTTCCAGATATTAATCTT 1560
DB 1501 TCTAGCGTTCGTCTCATCCACAAATAAGTTGTTGATTCGAGCTTCCAGATATTAATCTT 1560

QY 1561 TTTTAGATGGGTCTATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
DB 1561 TTTTAGATGGGTCTATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620

QY 1621 CATATAGCTCTGTTTGTGTTGTTGAGCTCTGGCTCTTTTGAAGACCCCTCTGCTTAATGTT 1680
DB 1621 CATATAGCTCTGTTTGTGTTGTTGAGCTCTGGCTCTTTTGAAGACCCCTCTGCTTAATGTT 1680

QY 1681 TTGTTGACCTTACAGCAATCCATATACGTTACGTTAGTGAAGTTCGACCCAAATGGTC 1740
DB 1681 TTGTTGACCTTACAGCAATCCATATACGTTACGTTAGTGAAGTTCGACCCAAATGGTC 1740

QY 1741 CAAATATAATTTAAATTTGGCCCAAAACAACATTTTACAAAAAATTTCAACAAACATGC 1800
DB 1741 CAAATATAATTTAAATTTGGCCCAAAACAACATTTTACAAAAAATTTCAACAAACATGC 1800

QY 1801 ATCGTTTCAAATTTTATTTTCAATGGCGTTATTTGTTTCAATTTGTAATATTTCTGTTTAA 1860
DB 1801 ATCGTTTCAAATTTTATTTTCAATGGCGTTATTTGTTTCAATTTGTAATATTTCTGTTTAA 1860

QY 1861 CTCACATGACCAATTTTATTTTCAAGAGAAACATTTTGTGATATAAATTAACATTTTA 1920
DB 1861 CTCACATGACCAATTTTATTTTCAAGAGAAACATTTTGTGATATAAATTAACATTTTA 1920

QY 1921 TGGAAACCAACCGTTAAGTCGATGATTTTGTAGTTTTAGTTTTGTCGTTTTGTGAATCAT 1980

Db 1530 TCTAGCGTCTGTCATCCCAATAGATGTTGTTGATCGAGCTTCCAGATTAATAATCTT 1589
 QY 1561 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCCATATAATTTCTAA 1620
 Db 1590 TTTTAGATGGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCCATATAATTTCTAA 1649
 QY 1621 CATATACGTCCTTTGTTTGGTAGGCTCTGCGTCTTTGAGACACCCCTTCCTAATGTT 1680
 Db 1650 CATATACGTCCTTTGTTTGGTAGGCTCTGCGTCTTTGAGACACCCCTTCCTAATGTT 1709
 QY 1681 TTGTTGACCTTAGACAATCCATAATAGTTCAGTTCGAGTTCGACCAAAATGTC 1740
 Db 1710 TTGTTGACCTTAGACAATCCATAATAGTTCAGTTCGAGTTCGACCAAAATGTC 1769
 QY 1741 CAATATATATTTAAATTTGGCCACAAAACAAATTTTACAAACAAATTTCAACAAATGTC 1800
 Db 1770 CAATATATATTTAAATTTGGCCACAAAACAAATTTTACAAACAAATTTCAACAAATGTC 1829
 QY 1801 ATCGTTTCAAAATTTTATTTATTCATGCGGTATTTGTTCAATGTAATAATTTCTGTTTAA 1860
 Db 1830 ATCGTTTCAAAATTTTATTTATTCATGCGGTATTTGTTCAATGTAATAATTTCTGTTTAA 1889
 QY 1861 CTCACTGACGAATTTTAAATTTTCAAGAGACATTTTGTATATAATAACATTTTA 1920
 Db 1890 CTCACTGACGAATTTTAAATTTTCAAGAGACATTTTGTATATAATAACATTTTA 1949
 QY 1921 TGAACACACCGGTTAAGCTCGATGATTTTGTAGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1980
 Db 1950 TGAACACACCGGTTAAGCTCGATGATTTTGTAGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 2009
 QY 1981 TAAGACCTACATTTGATCCCTCATTTCTTAATTAATTAAGAAATCAACATGATTA 2040
 Db 2010 TAAGACCTACATTTGATCCCTCATTTCTTAATTAATTAAGAAATCAACATGATTA 2069
 QY 2041 GTTCACCAAGACGCTCTTATGCTATTAAGAGTTCAGACCAAGGATGACCGGGTCAT 2100
 Db 2070 GTTCACCAAGACGCTCTTATGCTATTAAGAGTTCAGACCAAGGATGACCGGGTCAT 2129
 QY 2101 TAAGACCTCTATATTCACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2149
 Db 2130 TAAGACCTCTATATTCACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2178

RESULT 4
 US-10-027-880-3
 ; Sequence 3, Application US/10027880
 ; Publication No. US20030106105A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HOFFMAN, BEATE
 ; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
 ; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
 ; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
 ; FILE REFERENCE: CHEP:00303
 ; CURRENT APPLICATION NUMBER: US/10/027,880
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: PCT/FR00/01768
 ; PRIOR FILING DATE: 2000-06-23
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 4413
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-10-027-880-3
 Query Match 100.0%; Score 2149; DB 15; Length 4413;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGAATTTGTATATTTGAAGCAATCTCAAAAGATAAGTGGGATATATAACACCG 60
 Db 136 GTCGAATTTGTATATTTGAAGCAATCTCAAAAGATAAGTGGGATATATAACACCG 195

QY 61 GCGAAGTACAGTCTTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 120
 Db 136 GCGAAGTACAGTCTTACCTTTTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 255
 QY 121 AATTCCTGAATCTTTCAITTTCTGAATTTGATATTTTACATTTTATCAAAAAAAGTACA 180
 Db 256 AATTCCTGAATCTTTCAITTTCTGAATTTGATATTTTACATTTTATCAAAAAAAGTACA 315
 QY 181 AGTTCTCAAAAGACAGAGTTAAACAACTTGTGTGTCAAATGCTTAATTTAAAGCCTAA 240
 Db 316 AGTTCTCAAAAGACAGAGTTAAACAACTTGTGTGTCAAATGCTTAATTTAAAGCCTAA 375
 QY 241 TCTATGATTTCCCTTTTCTTACGATATATCTGATATGATATGATATGACCCCATTTGTTG 300
 Db 376 TCTATGATTTCCCTTTTCTTACGATATATCTGATATGATATGACCCCATTTGTTG 435
 QY 301 TCATTAACCTTCCACTCTATATCATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
 Db 436 TCATTAACCTTCCACTCTATATCATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
 QY 361 GTGATATATTTGTAAGAAAAAAGTGGTATCTGCTATATCAATACACCGTC 420
 Db 496 GTGATATATTTGTAAGAAAAAAGTGGTATCTGCTATATCAATACACCGTC 555
 QY 421 TCGAATTCCTCAACAAATTTCTAGGAGAAATGACGTCCTCTTGGTTTATTTTATTT 480
 Db 556 TCGAATTCCTCAACAAATTTCTAGGAGAAATGACGTCCTCTTGGTTTATTTTATTT 615
 QY 481 CTTAATAACATCTCTATATTTTAAACACTTCGATGTCCTGCTTAAATTTTGAATGTC 540
 Db 616 CTTAATAACATCTCTATATTTTAAACACTTCGATGTCCTGCTTAAATTTTGAATGTC 675
 QY 541 TAAATTTCTTAATCATATAATCTGAAGAAATTCGTCGAAGACACAGGACATGCTAG 600
 Db 676 TAAATTTCTTAATCATATAATCTGAAGAAATTCGTCGAAGACACAGGACATGCTAG 735
 QY 601 GGCACGTAGTTTACCTTTAAAGAACCATCAAAATATATTAAAGAAAGAACTTCCCTAAA 660
 Db 736 GGCACGTAGTTTACCTTTAAAGAACCATCAAAATATATTAAAGAAAGAACTTCCCTAAA 795
 QY 661 AGAACAAATTTAATAAGTGGATATAAAGATAGAGGTAGGACAGAGAAAGCTATGG 720
 Db 796 AGAACAAATTTAATAAGTGGATATAAAGATAGAGGTAGGACAGAGAAAGCTATGG 855
 QY 721 CCGGACTCTGAACAGGACGTCCTCCACCATCTCGGAGACGCGGACGCTGATGATT 780
 Db 856 CCGGACTCTGAACAGGACGTCCTCCACCATCTCGGAGACGCGGACGCTGATGATT 915
 QY 781 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTTACAGGGTCACAAACCATATCC 840
 Db 916 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCTGCTTACAGGGTCACAAACCATATCC 975
 QY 841 AATGTTCTGCTTATTTATTAATACTAAAGATCCCTCTTGCTTTTGTCTTTTATTCGT 900
 Db 976 AATGTTCTGCTTATTTATTAATACTAAAGATCCCTCTTGCTTTTGTCTTTTATTCGT 1035
 QY 901 GATATATAATCTAACTTAATTTAGTTTCTAAATAATATATGCTCTACCTATGTTTCTACTG 960
 Db 1036 GATATATAATCTAACTTAATTTAGTTTCTAAATAATATATGCTCTACCTATGTTTCTACTG 1095
 QY 961 ACCTCAGTCCCTAGTTAGCTTATATGACATATGTAAGAAATGACGCCCAAAATTTGAAGAG 1020
 Db 1096 ACCTCAGTCCCTAGTTAGCTTATATGACATATGTAAGAAATGACGCCCAAAATTTGAAGAG 1155
 QY 1021 TTCTCTTCTGCAACTAACTCTTATCTTACTCTATGAGCTATGTTTAAATATTTGAATGTT 1080
 Db 1156 TTCTCTTCTGCAACTAACTCTTATCTTACTCTATGAGCTATGTTTAAATATTTGAATGTT 1215
 QY 1081 GGCACCTCTGATTAATATATGACGTTGACCTGATATAAAGAAATGATAGACATTTAG 1140
 Db 1216 GGCACCTCTGATTAATATATGACGTTGACCTGATATAAAGAAATGATAGACATTTAG 1275


```
RESULT 6
US-10-021-323-17801/c
; Sequence 17801, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; CURRENT APPLICATION NUMBER: US-10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 17801
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(396)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3831-002-Q1-N6-A7
US-10-021-323-17801

Query Match          2.8%; Score 60.6; DB 17; Length 396;
Best Local Similarity 46.8%; Pred. No. 0.021;
Matches 159; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 374 AAAAAAAAAAAAAAGTGATATCTGCTATATACAAATACCAACGCTCGAATTGCCCTCA 433
Db |||||
QY 391 AAAATATATAAATTATTNTTATTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 332
Db |||||
QY 434 ACATTTCTAGAGAAATGACGTGCTCTTTGGTTTATTTTATTTCTTAATACATAC 493
Db |||||
QY 494 TCTATATTTTAAACATTCGATGCTCGCTAAATTCGATGCGCTAAATTTCTCTAA 553
Db |||||
QY 554 TCATAAATCGTAAAGAAAATTCGTCGAGCCACAGGACATGCATAGGCGAGTAGTAC 613
Db |||||
QY 614 CTTTAAACCATCAAAAATATATTAATAGAAAAGGAACCTTCCTPAAAAGAACATTTAAT 673
Db |||||
QY 674 AAAGTGATATAAAAAAGATAGAGGTAGGACAGAGAAA 713
Db |||||
QY 91 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 52

RESULT 7
US-10-311-455-1280
; Sequence 1280, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1280

Query Match          2.8%; Score 59.8; DB 15; Length 7597;
Best Local Similarity 42.7%; Pred. No. 0.13;
Matches 429; Conservative 0; Mismatches 567; Indels 9; Gaps 2;

QY 1025 TCTTCTGCAACTAACTCTTATCTTACTCATTTAGCTATGTTAAATTTGAATGTTGGCA 1084
Db |||||
```

```
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1280
; LENGTH: 6175
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1280

Query Match          2.8%; Score 60.2; DB 15; Length 6175;
Best Local Similarity 50.2%; Pred. No. 0.095;
Matches 149; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 1113 TAGATAAAAAACATGATAGACATTTAGTTTAAACTTGAAATGTTTGAACCTTTG 1172
Db |||||
QY 1173 GATTACGTGGATTGTTGATGGAATPAAATTTTGAAGATATTTATATATTGAAGATGTTTA 1232
Db |||||
QY 1233 TATATATTAGAGTTTATATAGCAGAAAATATTTGATAGATGTTGCTCTTTGTAGTTAC 1292
Db |||||
QY 1293 TCTTTTGTGCGTAGTCCTTTCTCCCTCATCCCTATGAGAAAAATCCAAATAGTTT 1352
Db |||||
QY 1353 AAGAAATTTTGTGTAATTCATAGTCCTTTTCGTAACACAGTTCTATGAGCTAT 1409
Db |||||
QY 5801 ATATTAGTTTTTTATATATTTTATTGATATTTTATTGATATTTTATTATATTTATATAT 5957
Db |||||

RESULT 8
US-10-311-455-986
; Sequence 986, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

Query Match          2.8%; Score 59.8; DB 15; Length 7597;
Best Local Similarity 42.7%; Pred. No. 0.13;
Matches 429; Conservative 0; Mismatches 567; Indels 9; Gaps 2;

QY 1025 TCTTCTGCAACTAACTCTTATCTTACTCATTTAGCTATGTTAAATTTGAATGTTGGCA 1084
Db |||||
```


Query Match 2.8%; Score 59.4; DB 17; Length 33053;
Best Local Similarity 51.1%; Pred. No. 0.31;
Matches 166; Conservative 0; Mismatches 156; Indels 3; Gaps 1;

QY 1064 GTTAATAATTGAATGCTGCACCTCTCGTATTAATAATATGCCAGTTCGACCTAGATAAAAA 1123
DB 8962 GTGAAAAATTAATAATAGTTTATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 9021

QY 1124 ACATGATAGACATTTAGTTTAAAACTTGAAATGTTTATTTGAACCTCTTTGGATTACGTGGA 1183
DB 9022 ATATAATTTATTTTCTTAATTAGAGAAATGTTTATTAATAATTAATAATTAATAATTAATA 9081

QY 1184 TTGTTGATGGAATTAATTTTGAAGATTTATATATTAATTAATTAATTAATTAATTAATTAATA 1243
DB 9082 TTTGATTTGAGTAAGTATTTTGTGAAATTTGGTTTTGG--TGTTTTTTTTTTT 9138

QY 1244 GTTTATATAGCAGAAATATTAATGATAGATGTTGTCCTTTTGTAGTACTCTTTTGTGTT 1303
DB 9139 TATTAAATGATAAAAAATTTTATTTATTTGTAATTAATTTTGTTCGTCGGGTTTTTATC 9198

QY 1304 GCGTAGTCTTTCTCTCATCTCTCTATGAAAGAAATCCAAATAGTTTAAAGGAAATTTT 1363
DB 9199 GGTAATAATTTTATTTAATAGTAGTAAATAATAGTATTTTGAAAGTATAGCGATTAAT 9258

QY 1364 TGTGTAATCAATAGCTTTTTCGTA 1388
DB 9259 TTAGTTTTTGTATATTTTGTGTTA 9283

RESULT 11

US-10-221-714A-418
; Sequence 418, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 418
; LENGTH: 5413
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-418

Query Match 2.7%; Score 58.6; DB 13; Length 5413;
Best Local Similarity 44.0%; Pred. No. 0.19;
Matches 351; Conservative 0; Mismatches 434; Indels 12; Gaps 2;

QY 1135 ATTTAGTTTAAACTTGAAATGTTTATTTGAACCTTTGGATTAGCTGATTTGTTGTTATGG 1194
DB 1249 ATGTAATTTAGATTTTATTAATTAAGATGAATTTATTTTGTAGTTTATTTTGTATAT 1308

QY 1195 ATTAATTTTGAAGATATTTATATATTAATTAAGATGTTTATATATAT-TAGAGTTTATATAG 1253

DB 1309 ATATATTTTATATATATATATTTATATATATTTTATATATATATATATATATTTTATATAT 1368
QY 1254 CAGAAAAATTAATGATAGATGTTGCTCTTTTGTAGTTACTCTTTTGTGTTGGTAGTCTCT 1313
DB 1369 ATATATTTTATATATATATATAGAGGTATTTTATTTATTTTATTTTATAGAGGTATTTAT 1428
QY 1314 TTCTCTCATCTCTCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTCTGTTGTAATTC 1373
DB 1429 ATATATATATTTTATATATAATATATATTTATTTTATAGGTATTTTATATATATTT 1488
QY 1374 A-----TAGTCTTTTCCGTAACCAACAGTTCTATGTAGCTATCGTCAATCATATTC 1422
DB 1489 TATATATAGATTTTATTTTATATATAGGTATTTATATTTATTTAGTAATATATATAT 1548
QY 1423 CTCTTTGCAACACAAAAAGATCGTTTGTGAAAATTTAGTGGGCTCTAAAGTCGCA 1482
DB 1549 ATTTATTTTATATAGTGTGTTTTTATATATATATTTATTTTATATATAGGTATTTA 1608
QY 1483 TTTGTTGCTCTCTCGAAATCTAGCGTTCTGTCATCCACAAATAGTTCGTTGATTCGAGC 1542
DB 1609 TTTTATATACGTATTTATATATATATTTATGTATATATTTTATATATTTTATATAT 1668
QY 1543 TTCCAAGATTAATCTTTTGTAGATGGGTCTGATGAAGATTTCTAACTTCGTATACGAGTG 1602
DB 1669 TATATAGTTTTTATATTTAGGAATATATATATTCGTTTATATATTTATATATAGTTATTT 1728
QY 1603 TATCCATATAATTTCTAACATATAGCTCTGTTTGTGTTGGTAGGCTCTGCGTCTTTGAGAC 1662
DB 1729 TATATATATATATATACGTATTTATTTTATATATATACGTATTTTATATATATATAT 1788
QY 1663 CACCCCTTGTCTAATGTTTGTGTCACCTTAGCAATCCATAATACGTTACGTGAGTCGA 1722
DB 1789 TTTTATTTTATTAATATATTTATATATATATATATATAGGTATTTTATATATTTATATA 1848
QY 1723 AGTTGCCACCAATGGTCCAAATATAATTTAAATTTGGCCACAAACAAACATTTTACAAA 1782
DB 1849 TTTATGTATATATTTATATATTTTATATATATATATATATAGGTATTTTATATAGATA 1908
QY 1783 CAAATTCACAAACATGCTGCTTCAATTTTATTTATTCATGGGCTTAATTTGTTTAT 1842
DB 1909 TTTTATATATTTATATATTTATATATATAATAATTTATATATAGGTATTTTATATAT 1968
QY 1843 TGTAAATATCTGTTTAACTCAGTCAGCAATTTTAAATTTTCAAGAGAAACATTTT 1902
DB 1969 TTTATATATATAGTTTGTGTTTGAATTTTGTGTTTATTTTAAATTTTATTTATTTAT 2028
QY 1903 GATATAAATAACATTTT 1919
DB 2029 TTTTATTTAATGTTTT 2045

RESULT 12

US-10-433-793-28
; Sequence 28, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 28
; LENGTH: 37515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-28

Query Match 2.7%; Score 57.6; DB 17; Length 37515;
Best Local Similarity 48.8%; Pred. No. 0.77;

Matches	156;	Conservative	0;	Mismatches	164;	Indels	0;	Gaps	0;
QY	1146	AACTTGAATGTTATTTGAACCTCTTTGGATTACGTGGATTGTTGTATGGATTAAATTTTG	1205						
Db	6823	AAGTATAAGTTTITAGAGATATTTTGTAGTTTTTTTGTGTAGTAAAGTTAATAAAATTTTA	6882						
QY	1206	AGATATTTTATATTTGAAGATGTTTATATATATTAGAGTTTATATACGACGAAAAATATTC	1265						
Db	6883	ATAATATTTAAAGGATGTTTATGGTTATTTTGGTAAAGAAATAATTATATTGATAAAGA	6942						
QY	1266	ATGTAGATGTTGCTTTTTGTAGTTACTCTTTTTCGTTGCGTAGTCTTTTCCTCCTCATCC	1325						
Db	6943	TTGTAGTATTTGATATTTAGAAATAATATTTTGGATTTTTTTTGTATTTTAAATAAAA	7002						
QY	1326	TCCTATGAAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAAATCATAGTCTTTTTC	1385						
Db	7003	TGATTTGTGAGAAATGTAATTTTTTTTTTAGGTAATATTTTTTTTTAAATGTTATTTAT	7062						
QY	1386	GTAACACACAGTCTATGTAGTACTATCGTCATCATATCTCTTTGCAACAACAAAAAGAT	1445						
Db	7063	ATAAAATTAATTAATTTGAAATATCGTATAAAGTTATTTTTTTTAAAAATAAAAAAGT	7122						
QY	1446	CGTTTTTTGTAATAATTTAGTA	1465						
Db	7123	AAATATTTGTTGTAATAATA	7142						
RESULT 13									
US-10-311-455-1301									
; Sequence 1301, Application US/10311455									
; Publication No. US20030143606A1									
; GENERAL INFORMATION:									
; APPLICANT: OLEK, Alexander									
; APPLICANT: PIEPENBROCK, Christian									
; APPLICANT: BERLIN, Kurt									
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System									
; TITLE OF INVENTION: cytosine methylation									
; FILE REFERENCE: 5013.1014									
; CURRENT APPLICATION NUMBER: US/10/311,455									
; CURRENT FILING DATE: 2002-12-16									
; PRIOR APPLICATION NUMBER: PCT/EP01/07537									
; PRIOR FILING DATE: 2001-07-02									
; PRIOR APPLICATION NUMBER: DE 10032529.7									
; PRIOR FILING DATE: 2000-06-30									
; PRIOR APPLICATION NUMBER: DE 10043826.1									
; PRIOR FILING DATE: 2000-09-01									
; NUMBER OF SEQ ID NOS: 2424									
; SEQ ID NO 1301									
; LENGTH: 8277									
; TYPE: DNA									
; ORGANISM: Artificial Sequence									
; FEATURE:									
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)									
US-10-311-455-1301									
Query Match 2.6%; Score 56.8; DB 15; Length 8277;									
Best Local Similarity 53.1%; Pred. No. 0.54;									
Matches 121; Conservative 0; Mismatches 107; Indels 0; Gaps 0;									
QY	1096	AAATATGCCAGTTGCACCTAGATAAAAAACATAGACATTTTAGTTTAAAACTTGAAT	1155						
Db	6529	AAATATCGTGTAAATATAGGAATATATATGATAAAATATTATTATTATTATATAGAT	6588						
QY	1156	GTTATTTGAACCTTTGGATTACGTGGATGTTCTGTATGGATTAATTTTGAAGATATTTA	1215						
Db	6589	TTTATTGAGGTTTTTTCGATATTTTGTTTGTTTTTGTAGTAAAAATTTTTTTTAGGTT	6648						
QY	1216	TATATTGAAGATGTTTATATATATTAGGTTTATATAGCAAAAATATTGATGTAGATGT	1275						
Db	6649	TAGGATTAGTTTAAGGATTAATGTTATATAGTTTTTAAATTTTTTTTAAATTTGGATTAGTTT	6708						
QY	1276	TGTCCTTTTTGTAGTTACTCTTTTTTTGTTGCGTAGTCCTTTCTCCTCAT	1323						

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 04:03:48 ; Search time 5558 Seconds

(without alignments)

11546.208 Million cell updates/sec

Title: US-10-027-880-1

Perfect score: 2149

Sequence: 1 gtcgaattgtgatattgt.....actaatgtctaataatcag 2149

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	355	16.5	355	29	CC793681 SALK 0173
2	159	7.4	833	28	BZ457686 B0NRW80TR
3	114.4	5.3	703	28	BZ470713 B0NNA88TR
4	98.8	4.6	690	28	BH959481 odd91g04.

c	5	87.2	4.1	92	29	BX285678	Arabidops
c	6	85.2	4.0	1200	13	BX437758	BX437758
c	7	80	3.7	1101	29	CNS0039G	AL063921 Drosophil
c	8	78.6	3.7	646	28	BH963752	BH963752 odf98b01.
c	9	76	3.5	1201	9	AL565455	AL565455
c	10	73.6	3.4	1101	29	CNS00EVL	AL069706 Drosophil
c	11	73	3.4	928	29	CNS00DKY	AL071865 Drosophil
c	12	71.4	3.3	379	28	BH757746	BH757746 SALK 0081
c	13	70.8	3.3	604	28	AQ964121	AQ964121 LERG742TR
c	14	70.8	3.3	697	28	AQ964120	AQ964120 LERG742TF
c	15	70.6	3.3	1200	13	BX437758	BX437758
c	16	69.8	3.2	1896	29	CG753083	CG753083 P048-1-C0
c	17	69.4	3.2	1101	29	CNS00EO7	AL069440 Drosophil
c	18	69.4	3.2	1200	13	BX415878	BX415878
c	19	69.2	3.2	726	28	B26447	B26447 F2K20TF IGF
c	20	68.4	3.2	1200	13	BX415878	BX415878
c	21	67.8	3.2	1101	29	CNS0039G	AL063921 Drosophil
c	22	67.8	3.2	1201	13	BX439779	BX439779
c	23	67.2	3.1	961	14	CK230311	CK230311 ILDMUTGEN
c	24	66.2	3.1	83	29	AL763344	AL763344 Arabidops
c	25	66	3.1	1201	9	AL536104	AL536104
c	26	66	3.1	1201	13	BX360615	BX360615
c	27	65.8	3.1	956	13	BX414969	BX414969
c	28	65.6	3.1	1165	13	BX338369	BX338369
c	29	65.2	3.0	1201	13	BX443774	BX443774
c	30	65	3.0	1201	13	BX462207	BX462207
c	31	64.8	3.0	994	13	BX414650	BX414650
c	32	64.6	3.0	701	28	AQ250248	AQ250248 T3F21-Sp6
c	33	64.6	3.0	1056	13	BX415058	BX415058
c	34	64.6	3.0	1206	28	BZ695529	BZ695529 SP_Ba006
c	35	64.4	3.0	1101	29	CNS0039R	AL063932 Drosophil
c	36	64.2	3.0	1101	29	CNS002M0	AL062875 Drosophil
c	37	64.2	3.0	1101	29	CNS00EVL	AL069706 Drosophil
c	38	64.2	3.0	1392	29	CG757503	CG757503 P052-4-C0
c	39	64	3.0	1167	29	CNS07360	AL427102 clone BA0
c	40	64	3.0	1201	13	BX461824	BX461824
c	41	63.8	3.0	1200	13	BX437739	BX437739
c	42	63.8	3.0	1201	9	AL536104	AL536104
c	43	63.4	3.0	1029	29	CNS01ZGM	AL174271 Tetraodon
c	44	63.4	3.0	1098	13	BX377526	BX377526
c	45	63.2	2.9	1101	29	CNS016LI	AL106896 Drosophil

ALIGNMENTS

RESULT 1

CC793681

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK_017399.55.00.x, genomic

survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.

1 (bases 1 to 355)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,

Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,

Shinn,P., Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salik.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

FEATURES
source
1. .355
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_017399.55.00.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 16.5%; Score 355; DB 29; Length 355;
Best Local Similarity 100.0%; Pred. No. 5.4e-51;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1790 AACAAACATGCATCGTTTCAAAATTTTATTTCAATGGCGTTATTTGTTTCATGTAAAT 1849
Db 1 AACAAACATGCATCGTTTCAAAATTTTATTTCAATGGCGTTATTTGTTTCATGTAAAT 60
QY 1850 ATTCTGTTTAACCTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTTCATATAA 1909
Db 61 ATTCTGTTTAACCTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTTCATATAA 120
QY 1910 ATAACTTTTATGAACACCGGTTAAGCTCGATGATTTTGAGTTTGTCTGTTT 1969
Db 121 ATAACTTTTATGAACACCGGTTAAGCTCGATGATTTTGAGTTTGTCTGTTT 180
QY 1970 TGTGAATCATTAAGACCTACATTTGATCCCTCATTTTAAATTTAGGAATCAAC 2029
Db 181 TGTGAATCATTAAGACCTACATTTGATCCCTCATTTTAAATTTAGGAATCAAC 240
QY 2030 ATGATGATTAAGTTTCAACCAAGACGCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATG 2089
Db 241 ATGATGATTAAGTTTCAACCAAGACGCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATG 300
QY 2090 ACCGGGTCTATTAAGACGCTCTTATTAACACCTACTCCTCAATTTGCTAAATTA 2144
Db 301 ACCGGGTCTATTAAGACGCTCTTATTAACACCTACTCCTCAATTTGCTAAATTA 355

RESULT 2
BZ457686/c
LOCUS
DEFINITION
BONRW80TR BO.1.6.2 KB tot Brassica oleracea genomic clone BONRW80,
genomic survey sequence.
ACCESSION
BZ457686
VERSION
BZ457686.1 GI:26737311
KEYWORDS
GSS.
SOURCE
Brassica oleracea
ORGANISM
Rukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 833)
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BONRW80TF
Contact: Chris Town
TIGR

712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: Sheared ends.

FEATURES
source
1. .833
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONRW80"
/clone_lib="BO.1.6.2 KB tot"
/note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match 7.4%; Score 159; DB 28; Length 833;
Best Local Similarity 62.6%; Pred. No. 1.2e-17;
Matches 420; Conservative 0; Mismatches 200; Indels 51; Gaps 9;
QY 584 CACAGGACATGATAGGGCAGCTAGTTTACCTTTAAACCATCAAAA-TATATTATATAG 642
Db 818 CAAAGGACATGATGGGCGACGTAGTTACTTTAGACCATAAAAAATTAACATAATAG 759
QY 643 AAAAGGAACTTCTTAAAGAAACAATT-----TAATAAGTGGATAAAAAAGATAAG 695
Db 758 AGAAGGAACTTCTTAAATTAAGATAAATAAATAAAGAAAGAAAGAAACAAGAA 699
QY 696 AAGTAGGCGAAGAAACATATGGCGGACGTCGTAAACAAAGGAGCTCCGACCATGTC 755
Db 698 GGTATAGGCAAAAGAAACATATAGCGGACCGTAACAAAGGAGCTCTCGACCATGTC 639
QY 756 GGAGACGGCAGAGCGCTGACTGATTTTCT-----TTTTCTTTTCTTAAGACGTTGTTT 812
Db 638 GGAGACGGCAGAGCGCTGACTGATTTTCTTTTATCTTTTCTTCTTAAGACGTTGTTT 579
QY 813 CGTGCTTACAAAGGCTCAAAACCATATCCAAATGTTCTGCGCTATTATTATATACTAAAGA 872
Db 578 CGTGCTCAAAAGGCTCA-----GTGTCAACCATATCAAGATATAAA 536
QY 873 TCCCTCTTGTGCTTTGCTTTTATTCGTGATATAATACTAACTTAATTTAGTTCTAAAA 932
Db 535 CCGCTTCTGTAGTTTGTCTTTACTCGTGATAAAT-CTAAGTTTGAATTTAGTTCTAAT 477
QY 933 TATATATGCTCCTACCTATGTTTCTACTGACCTCAGTCCCTAGTTAGCTATATGACATAT 992
Db 476 TCAAAATTAATATGCA-----GTGACTCTATTTCCTCGGTCCTT 433
QY 993 GTGAAATGACGCCCAAAATTTGAAGAGTTTCTCTTCTCTGCG-----AACTAACTCTTATCTT 1049
Db 432 GTGACATGAGCTGGAATTTGAAGAGCTAGTTGTTCTACCTGCAACAAATTTATCTT 373
QY 1050 ACTCATGAGCTATGTTTAAATATATGATGTTGGCAGCTCTCG--TATTAAATATGCCAGTT 1107
Db 372 AAGCATGTTTCAAGTGTAAAGTACTGTAAGAGCTGAGCTTTGTATATATAATATGGCATCT 313
QY 1108 GCACCTAGATAAAAAACATGATAGACATTTAGTTTAAAACTTGAATGTTTATTGAACT 1167
Db 312 GCACCTACCAAAACCATGATAATTTTAACTTATTATTTTCTAACAGGATTTGTC- 254
QY 1168 CTTTGGATTACGTTGATGTTTGTATGATTAATTTTGAAGATATTTATATATGAAGAT 1227
Db 253 GTATGAATTTTAAAGATGATCGTATACGTAACGTAATAAAGCGTATATAGATATCCAGAT 194
QY 1228 GTTATATATA 1238
Db 193 CCATATATTTA 183

RESULT 3
BZ470713
LOCUS
DEFINITION
BONNA88TR BO.1.6.2 KB tot Brassica oleracea genomic clone BONNA88,
genomic survey sequence.
703 bp DNA linear. GSS 13-DEC-2002

```
ACCESSION      BZ470713
VERSION        BZ470713.1  GI:26768062
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
TITLE          1 (bases 1 to 703)
JOURNAL        Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
COMMENT        Whole genome shotgun sequencing of Brassica oleracea
                Unpublished (2001)
                Other GSSs: BONNA88TF
                Contact: Chris Town
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TR
                Class: Sheared ends.
FEATURES       Location/Qualifiers
                source
                1..703
                /organism="Brassica oleracea"
                /mol_type="genomic DNA"
                /strain="T01000DH3"
                /db_xref="taxon:3712"
                /clone="BONNA88"
                /note="Vector: BO_1.6-2_KB tot"
                /note="Vector: PHOS1; Site_1: BstXI; 1.6-2 kb sheared
                total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN
Query Match      5.3%; Score 114.4; DB 28; Length 703;
Best Local Similarity 77.6%; Pred. No. 5.4e-10;
Matches 166; Conservative 0; Mismatches 41; Indels 7; Gaps 2;
QY 584 CACGGGACATCATAGGCGAGTGTACCTTTAAACCATCAAAAA-TATATTAATAG 642
      |||
      |||
      |||
Db 490 CAAAGGACATGTCGGCGCATGTACTTTAGACCATAAAAAATTAACCTAATAG 549
      |||
      |||
      |||
QY 643 AAAAGGAACTCTCTAAAGAAC-----AATTTAAAGTGGATAAAAAAGATAAGA 696
      |||
      |||
      |||
Db 550 AGAAGGAACTCTCTAAATAAGATAAAATAAAATAAAAGGAAAGAAAAACAAAGAG 609
      |||
      |||
      |||
QY 697 AGGTAGGAGAGAAACGTATGGCGGACCTCGTAACAAGGAGCGTCCCGACCACTGG 756
      |||
      |||
      |||
Db 610 GTATAGGCAAAAGAAACGTATAGCGGACCGCCGTAAACAAGGAGCGTCTGACCACTGG 669
      |||
      |||
      |||
QY 757 GAGACGGCGAGACGCTGACTGATTTTCTTTT 790
      |||
      |||
      |||
Db 670 GAGACGGCGAGACGCTGACTAATTTTATCT 703
      |||
      |||
      |||
RESULT 4
BH959481/c
LOCUS          BH959481
DEFINITION    odd91g04.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
                sequence.
ACCESSION      BH959481
VERSION        BH959481.1  GI:23440708
KEYWORDS       GSS.
SOURCE         Brassica oleracea
ORGANISM       Brassica oleracea
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
TITLE          1 (bases 1 to 690)
JOURNAL        Delchaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
                Nash, W., Rabinowicz, P.D. and Wilson, R.K.
COMMENT        Whole genome shotgun reads from Brassica oleracea
                Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odd91 row: g column: 04
Seq primer: -21UPOT forward
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 551.
FEATURES       Location/Qualifiers
                source
                1..690
                /organism="Brassica oleracea"
                /mol_type="genomic DNA"
                /db_xref="taxon:3712"
                /clone_lib="B.oleracea002"
                /note="Vector: pOTw13; Whole genome shotgun library from
                flowering buds. DNA was purified from a crude nuclear
                prep using Brassica oleracea T01000DH3 buds provided by
                Thomas Osborn at the University of Wisconsin. Genomic
                DNA was provided by Pablo Rabinowicz (CSHL) and the
                shotgun library prepared at Washington University Genome
                Sequencing Center."
ORIGIN
Query Match      4.6%; Score 98.8; DB 28; Length 690;
Best Local Similarity 84.2%; Pred. No. 2.5e-07;
Matches 123; Conservative 0; Mismatches 22; Indels 1; Gaps 1;
QY 2005 TTACTTTAATATAGGATCAACATCATGATTAAGTTCACCAAGAGCTCTCTATGG 2064
      |||
      |||
      |||
Db 551 TTTATTACTTTATTAACCAAGAAACAAGATGAAGTTTCATCAAGAGTCCCTATGG 492
      |||
      |||
      |||
QY 2065 CTATTAAAGAGTCAGACGCAAGGATGACCGGGTCATTAAAGAGCTTTATAT-TCACCAT 2123
      |||
      |||
      |||
Db 491 CTATTAAAGAGTCAGATGCAAGATGACCGGGTCATTAAAGATCTTATATCGAACCAT 432
      |||
      |||
      |||
QY 2124 TACTCCACTAATGCTAATTAATCAG 2149
      |||
      |||
      |||
Db 431 TACTCCACTAATTAATTAATTCAG 406
      |||
      |||
      |||
RESULT 5
BX285678/c
LOCUS          BX285678
DEFINITION    Arabidopsis thaliana T-DNA flanking sequence GK-384F07-017271,
                genomic survey sequence.
ACCESSION      BX285678
VERSION        BX285678.1  GI:28884674
KEYWORDS       GSS.
SOURCE         Arabidopsis thaliana (thale cress)
ORGANISM       Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1 Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.
                and Weisshaar, B.
                A pipeline for automated high-throughput generation of FSTs
                (flanking sequence tags) from Arabidopsis thaliana T-DNA
                transformed lines
                Unpublished
                2
                Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
                A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
                for flanking sequence tag based reverse genetics
                Unpublished
                3 (bases 1 to 92)
                Rosso, M., Strizhov, N., Li, Y. and Weisshaar, B.
                Direct Submission
                Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer
                Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
                This sequence is recovered from the left border of the T-DNA. It
                indicates an insertion within the locus defined by clone T32E8. The
```


sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

```

1..92
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-384F07-017271"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (Ti) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

```

ORIGIN

```

Query Match 4.1%; Score 87.2; DB 29; Length 92;
Best Local Similarity 96.7%; Pred. No. 7.6e-05;
Matches 89; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1873 TTTTAAATTTTCAAGAAGAACATTTTGATATAAATAACATTTTATGTAACACCGG 1932
DbbTTTTTAAATTTTCAAGAAGAACATTTTGATATAAATAACATTTTATGTAACACCGG 33

Qy 1933 TTAAGCTCGATGATTTTGGATTTTGGT 1964
DbbTTTTTAAATTTTCAAGAAGAACATTTTGATATAAATAACATTTTATGTAACACCGG 33

```

RESULT 6

BX437758/c

```

LOCUS BX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP008YB01
5-PRIME, mRNA sequence.

```

BX437758

BX437758.1 GI:30773605

EST.

Homo sapiens (human)

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

1 (bases 1 to 1200)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CSOCAP008CA01QP1.

Location/Qualifiers

1..1200

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSOCAP008YB01"

/tissue_type="THYMUS"

/clone_lib="Homo sapiens THYMUS"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, into double-strand cDNA was digested with Not I and cloned, into the Not I and EcoRV sites of the pCMVSPORT 6 vector."

FEATURES

source

```

1620 ACATATACGCTCTGTTTGGTAGGCTCTGGCTCTTTTGTAGACACCCCTT 1671
: : : : : : : : : : : : : : : : : : : : : : : : : :
375 WATAAAAWAWAWTAWTATCCYCCYATATATTTTWWAAAAAAAHYYT 324

```

RESULT 7

CNS0039G

LOCUS

DEFINITION

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence T83 end of BAC #
 BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit

Library was not normalized."

```

ORIGIN
Query Match 4.0%; Score 85.2; DB 13; Length 1200;
Best Local Similarity 31.2%; Pred. No. 4e-05;
Matches 278; Conservative 174; Mismatches 425; Indels 15; Gaps 2;

Qy 780 TTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCGTCTTACAGGGTCAAAACCAATC 839
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
1200 TTTTATYTTAAWTTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1141

Qy 840 CAATGTTCTGCTTATTAATAAAGTAAAGTCCCTCTGCTGCTTCTTTTCTTTTCG 899
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
1140 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1081

Qy 900 TGATATATACTAACTAAATAGTTCTAAATATATATGCTTCTTCTTCTTCTTCTTCT 959
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
1080 AAAAAAAWTTTATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1021

Qy 960 GACCTCAGTCCCTAGTATGCTATGACATATGTAAGTAAAGTAAAGTAAAGTAAAGT 1019
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
1020 WTAWATAWATAAAAHWWTTTTTWWAAAAAAAWTTTTTWWTATATATAAAAWATTTW 961

Qy 1020 GTTCTCTCTCTGCACTAACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1079
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
960 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 901

Qy 1080 TGGCACTCTCGTATTAATATGCGAGTTGCACTTAGATAAAACATGATAGACATTTA 1139
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
900 TTTTCTCMYTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 841

Qy 1140 GTTAAACCTTGAAAGCTTATTTCAACTCTTTGATGTTGATGTTGATGTTGATGTTA 1199
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
840 TTTTCTTWTWAAWATTTTCTTWWAAATTTTCTTWWAAATTTTCTTWWAAATTTTCTT 786

Qy 1200 ATTTTGAAGATATTTATATATGAGATGTTTATATATATATATATATATATATATAC 1259
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
785 TTTWAAAAAATAWATAAAAHCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 726

Qy 1260 ATATGATGATGATGTTGTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1319
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
725 AWTTCCTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 666

Qy 1320 TCATCTCTCTATGAAGAAATCCAAATAGCTTTTAAAGAAATTTTGTGTAATTCATAGTC 1379
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
665 AYTCTHYWAAAAAACAATAAAWAAAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAA 606

Qy 1380 TTTTCTGTAACACAGTTCTATGCTAGCTATGCTATGCTATGCTATGCTATGCTATGCT 1439
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
605 CTYMWATAAAWAAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 546

Qy 1440 AAAGATCGTTTGTGTAATTTAGTAGGGCACTAAAGTCGTCATCTTGTGTCCTGTCGAA 1499
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
545 AAAAAAATAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 486

Qy 1500 ATCTAGGCTCTGCTATCCACAATAAGTTGTTGATGCTGAGCTTCCAGATATATAATCT 1559
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
485 AWTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 436

Qy 1560 TTTTATAGATGGTCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATATATCTTA 1619
Dbb : : : : : : : : : : : : : : : : : : : : : : : :
435 WWTWTWTWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 376

Qy 1620 ACATATACGCTCTGTTTGGTAGGCTCTGGCTCTTTTGTAGACACCCCTT 1671
: : : : : : : : : : : : : : : : : : : : : : : :
375 WATAAAAWAWAWTAWTATCCYCCYATATATTTTWWAAAAAAAHYYT 324

```

```
fly), genomic survey sequence.
AL063921
AL063921.1 GI:4941778
GSS.
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN
Query Match 3.7%; Score 80; DB 29; Length 1101;
Best Local Similarity 17.5%; Pred. No. 0.00032;
Matches 121; Conservative 286; Mismatches 283; Indels 0; Gaps 0;
QY 12 ATATATTGTAGCAATCTGAAAGAAATAAGTCGGATATATAACAACCGCGAAGATACA 71
DDB 398 ATAAWAWWWTTTTTTTAWAAWAAATAATTTWAAWAAWAAATAATTTWAAWAAWAAW 457
QY 72 AGTCTACCTTTTGGGATGGAACCATGTTTTTAGGATTTACTTTGTAATTCCTGAAT 131
DDB 458 AWTAWTTTATWTAWAAAAAATAATTTTWTWTATTTTATTTTATTTTATTTTAAW 517
QY 132 CTTTCATTTCTGATTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 191
DDB 518 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 577
QY 192 AGCAGAGAGTTTAAACAACTGTGTCTCAATGCTAAATTTAAAGCTTATCTTATGAT 251
DDB 578 YTHYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTY 637
QY 252 CTTCTTTCTTCAGTATATCTGATATGATGATGATGATGATGATGATGATGATGATG 311
DDB 638 AYYYYTTCYHYHYHWHHHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 697
QY 312 CCACTATACATCAGTATCTCMAAGTCGAATAACATATCCTATCCTATGATGATG 371
DDB 698 CHCYHYHYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHTYHT 757
QY 372 TGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 431
DDB 758 YAWAHAAWAWWHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 817
QY 432 CAACAAATTTCTAGGAGAAATGACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 491
fly), genomic survey sequence.
818 YUHTAYTCWTWTHHHMMTWTHYHHTTWHHTTTHHAWMMHTTTCWMMWHTTWTWATHCW 877
QY 492 ACTCTATATTTTAAACACACTTCGATGTCGCTGCTTAAATTTGGAATGCTGCTAAATTCCTCT 551
DDB 878 ACCTMHHHHHHHHHHHMAACHNHTTHCMCHHHHCTCHHHHTMYHNTCHMMWHHWHW 937
QY 552 AATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATGCATAGGACACGATGTT 611
DDB 938 HHMMWATWTTTMTTMMWMMCCMHHCYHMMHMYMCHYUCTCTHATTHYHMYCTCY 997
QY 612 ACCTTTAAACCATCAAAATATATATAGAAAGAAAGAACTTCCTTAAAGACACATTTA 671
DDB 998 HYCTWHTYTWTAWTAHAHTTATTTWMMHWAHATWMMWMMWATAWACTCHHHTW 1057
QY 672 ATAAAGTGCATAAAAAAGATAAGGTA 701
DDB 1058 YHTCTWYHTYHTYHMMWAWMMWMMHWHMYA 1087
RESULT 8
BH963752/c
LOCUS
DEFINITION
BH963752
VERSION
BH963752.1 GI:23444978
KEYWORDS
GSS.
SOURCE
Brassica oleracea
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 646)
AUTHORS
Delehaunty, K., Rewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
TITLE
Whole genome shotgun reads from Brassica oleracea
JOURNAL
Unpublished (2002)
COMMENT
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odf98 row: b column: 01
Seq primer: -21UPPOT forward
Class: shotgun
High quality sequence start: 64
High quality sequence stop: 551.
FEATURES
source
1..646
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T0100DH3 buds, provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
ORIGIN
Query Match 3.7%; Score 78.6; DB 28; Length 646;
Best Local Similarity 61.9%; Pred. No. 0.00076;
Matches 300; Conservative 0; Mismatches 109; Indels 76; Gaps 8;
QY 596 CATAGGCGAGTAGTACTCTTTAAACCATCAAAATATATTAATAGAAAGAACTTC 655
DDB 646 CATGTGGCAGCTAGTACTCTTTTAGACCATTAATAAACTACTCCGAATAAAG 595
QY 656 CTAAGACACATTTAATAAGTGTATTAAGTGTATTAAGTGTATTAAGTGTATTAAG 715
DDB 594 --AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 537
```

```
QY 716 TATGCGCGGAGCTC--GTAACAAGGAGCGTCCCGACACACTCGCGGAGCGGCGAGCGCTG 773
D 536 TATAGCCGCGGAGCGGCTAACAAAGGAGCGTCCCGACACTCGCGGAGATGGCGAG----- 483
QY 774 ACTGATTTTTCTTTTCTTTCTTAAAGACGTTGTTTCGTGCTTCAAGGGTCAAAAC 833
D 482 -----ACGTTATTTCTGCTGCACAAAGGTCAGTG 453
QY 834 CATATCCAAATGTTCTGCTATTAATATATAAAGATCCCTCTTTGCTGCTTTGCTTT 893
D 452 CGAAGCGCTCTCTATTAATGATAATATATAAATCCCGG--CCCATCTTTGCTGTTGCTT 395
QY 894 TATTCGTCATATAAATCTAACTTAAATAGTTCTTAAATATATATATGCTTACCTATGTT 953
D 394 TATTCGTCATATAA--ATCGAAGTTGCTTCTTAAATATATATATATATATGCT--ATGTT 338
QY 954 TCTACTGACCTCAGTCCCTCTAGTTAGCTATATGACATATGGAATAAGACCCCAAAATT 1013
D 337 TCTACTGCTCGTCCCT-----TGATGAATGACGTCGAAT 297
QY 1014 TGAAGAG-----TTCTCTCTCTGCACTAACTTCTTACTCTTACTCATTGAGCTATGTTAAA 1069
D 296 TGAAGAGCTATTTGTCGCTCTCCAACTAACTATTTCTTAAACATTGATCAGCGCTTTT 237
QY 1070 TATTG 1074
D 236 TTTTG 232

RESULT 9
AL565455/c
LOCUS
DEFINITION AL565455 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
VERSION CS0DF005YO18 3-PRIME, mRNA sequence.
KEYWORDS AL565455
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12916848.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9232.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF005BH09NP1&cluster=9232.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DF005BH09NP1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF005YO18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
```

FEATURES

source

RESULT 10

CNS000EVL/c

LOCUS

DEFINITION

CNS000EVL

Drosophila melanogaster

genome survey sequence

fly, genomic survey sequence.

AL069706

AL069706.1

GI:4949849

GSS.

Drosophila melanogaster (fruit fly)

ORIGIN

Query Match 3.5%; Score 76; DB 9; Length 1201;

Best Local Similarity 33.4%; Pred. No. 0.0015;

Matches 269; Conservative 127; Mismatches 402; Indels 8; Gaps 2;

QY 860 ATATACTAAAGATCCCTCTTGTGCTTTGCTTTATTCGGATATATAAATCTAACTTAA 919

D 1201 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1142

QY 920 ATTAGTTCTAAATATATATGCTTACCTATGTTTCTTACAGCCTCAGTCCCTAGTAGC 979

D 1141 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1083

QY 980 TATATGACATATGTGAAATGACGCCCAAAATTTGAAGAGTTCTCTCTCTCGCAACTTAA 1039

D 1082 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1023

QY 1040 CTCTTATCTTACTCAATGAGCTATGTTTAAATATTTGAATGTTGGCAGCTCTCGTATTA 1099

D 1022 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 963

QY 1100 TGCCAGTTGCACTAGATAAAAAAATGATAGACATTTAGTTTAAACTTGAATGTTA 1159

D 962 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 903

QY 1160 TTTGAATCTTTGATTTGAGTATTTGTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1219

D 902 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 843

QY 1220 TTGAAGATGTTTATATATATTTAGATTTATATAGCAGAAATATTTGATGATGATGATGAT 1279

D 842 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 790

QY 1280 CTTTGTGATTTACTCTTTTGTGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1339

D 789 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 730

QY 1340 ATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCTCTTTTCTGTAACCACTCT 1399

D 729 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 670

QY 1400 ATCTAGCTACGTCATCATATTTCTCTTTCACACAAACAAAGATCGTTTGTGTAAT 1459

D 669 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 610

QY 1460 TTATGAGGCACTAAAGTCGTCATTTGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1519

D 609 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 550

QY 1520 CAAATAGTTTGTGATTCGAGCTTCCAGATTTATTAATCTTTTGTAGATGGTCAAG 1579

D 549 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 490

QY 1580 ATTCTTAACTTCGTATACGAGTCTATCCATATAATTTCTAATATATATATATATATATAT 1639

D 489 WTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 430

QY 1640 GTAGGCTCTGCGCTTTTTCGAGACAC 1665

D 429 ATAGGATATATACCTTCTACTTAC 404

CNS000EVL
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B3 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.AL069706
AL069706.1 GI:4949849
GSS.
Drosophila melanogaster (fruit fly)


```

FEATURES
  source
    1. .697
      /organism="Arabidopsis thaliana"
      /mol_type="genomic DNA"
      /strain="LANDSBERG ERECTA"
      /db_xref="taxon:3702"
      /clone="LERGT42"
      /clone_lib="LERG"
      /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
      sheared to 0.4-0.7 Kbp before ligation."
    Class: shotgun.
    Location/Qualifiers
      Seq primer: TF
      Email: atotigr.org
      Fax: 301 838 0208
      http://www.tigr.org/tdb/at/at.html

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 08:23:39 ; Search time 16500 Seconds
(without alignments)
11592.291 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatcacagtgataaat.....tttcattgggttaacagagtcca 4413

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.un.*
- 28: em.vi.*
- 29: em.vt.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4413	100.0	4413	6	AX063411	AX063411 Sequence
2	4413	100.0	4413	8	AF242314	AF242314 Arabidops
3	4052	91.8	82454	8	AC012193	AC012193 Arabidops
4	2154	48.8	4309	6	AX063412	AX063412 Sequence
5	2149	48.7	2149	6	AX063409	AX063409 Sequence
6	2149	48.7	4280	6	AX063410	AX063410 Sequence
7	1486	33.7	1741	8	AF242313	AF242313 Arabidops
8	202	4.6	317	8	ATH553603	AJ553603 Arabidops
9	143	3.2	168	8	ATH522466	AJ522466 Arabidops
10	69	1.6	465	8	BT010186	BT010186 Arabidops
11	68	1.5	295	11	AL773135	AL773135 Arabidops
12	51	1.2	43982	8	AC007289	AC007289 Arabidops
13	26	0.6	15046	6	AX347002	AX347002 Sequence
14	26	0.6	124550	8	AC144806	AC144806 Medicago
15	26	0.6	125992	2	AC142222	AC142222 Medicago
16	26	0.6	137040	2	EX322621	EX322621 Daulo rer
17	26	0.6	155937	2	AC016341	AC016341 Homo sapi
18	26	0.6	174741	2	AC024280	AC024280 Homo sapi
19	26	0.6	199200	9	AC010351	AC010351 Homo sapi
20	25	0.6	773	10	BC025131	BC025131 Mus muscu
21	25	0.6	1140	8	AF037457	AF037457 Fritillar
22	25	0.6	1777	8	AF179249	AF179249 Lycopersi
23	25	0.6	2033	6	BD190852	BD190852 Secreted
24	25	0.6	47573	3	AF030694	AF030694 Plasmodiu
25	25	0.6	125479	9	HSJ244F24	AL096865 Human DNA
26	25	0.6	145456	9	AL513366	AL513366 Human DNA
27	25	0.6	349751	3	PFMAL4P3	AL035476 Plasmodiu
28	24	0.5	646	11	BV065942	BV065942 S212P6179
29	24	0.5	7752	6	AX348656	AX348656 Sequence
30	24	0.5	8964	6	AX251977	AX251977 Sequence
31	24	0.5	8964	6	AX344369	AX344369 Sequence
32	24	0.5	8964	6	AX346437	AX346437 Sequence
33	24	0.5	8964	6	AX349014	AX349014 Sequence
34	24	0.5	87114	9	AC005353	AC005353 Homo sapi
35	24	0.5	91469	9	AL356963	AL356963 Human DNA
36	24	0.5	94605	9	AF140480S2	AF360695 Homo sapi
37	24	0.5	95851	9	HS1118M15	AL109964 Human DNA
38	24	0.5	109860	9	AC112695	AC112695 Homo sapi
39	24	0.5	110000	2	AC139250_4	Continuation (5 of
40	24	0.5	110000	2	PFMAL7P1_03	Continuation (4 of
41	24	0.5	115866	9	AL627422	AL627422 Human DNA
42	24	0.5	120994	9	AC008697	AC008697 Homo sapi
43	24	0.5	124949	9	AL358794	AL358794 Human DNA
44	24	0.5	125021	8	AC007504	AC007504 Arabidops
45	24	0.5	126467	9	AF487553	AF487553 Homo sapi

ALIGNMENTS

RESULT 1
AX063411
LOCUS AX063411
DEFINITION Sequence 3 from Patent WO0100833.
ACCESSION AX063411
VERSION AX063411.1
KEYWORDS GI:12541199
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Hoffmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,
linear PAT 24-JAN-2001

recombinant vectors and host cells comprising same and transgenic plants obtained

JOURNAL Patent: WO 0100833-A 3 04-JAN-2001;

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)

FEATURES Location/Qualifiers

source 1..4413

/organism="Arabidopsis thaliana"

/mol_type="unassigned DNA"

/db_xref="taxon:3702"

ORIGIN

Query Match 100.0%; Score 4413; DB 6; Length 4413;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AAGATCCACAGTGAATAAATAAGAACGGATTCCGTGATATGCAACTATATATATGAA	60
Db	1	AAGATCCACAGTGAATAAATAAGAACGGATTCCGTGATATGCAACTATATATGAA	60
QY	61	ATTGAATACCTCGAATTCGGTTCGTATCAAGATCGAATCTCTAAAAACATATCTCTA	120
Db	61	ATTGAATACCTCGAATTCGGTTCGTATCAAGATCGAATCTCTAAAAACATATCTCTA	120
QY	121	TAATAAATATCTGCAGTCGAATTCGTATATTTGAAGCAATCTGAAAGAAATAAGTGG	180
Db	121	TAATAAATATCTGCAGTCGAATTCGTATATTTGAAGCAATCTGAAAGAAATAAGTGG	180
QY	181	ATATATAAACACCGCGGAAAGTACAAAGTTCTACCTTTTTTGGCATGGAACCATGTTT	240
Db	181	ATATATAAACACCGCGGAAAGTACAAAGTTCTACCTTTTTTGGCATGGAACCATGTTT	240
QY	241	TAGGAATTAATCTTGTAAATCTTCTGAATCTTTTCAATCTTCTGAATTCGATATTTAT	300
Db	241	TAGGAATTAATCTTGTAAATCTTCTGAATCTTTTCAATCTTCTGAATTCGATATTTAT	300
QY	301	CAAAAAAAGTACAAAGTCTACCAAGGACAGAGTTAAACAACTTGTTGTCAAATGC	360
Db	301	CAAAAAAAGTACAAAGTCTACCAAGGACAGAGTTAAACAACTTGTTGTCAAATGC	360
QY	361	TAATTTAAAGCCCTAACTTATGATTTCCCTTTTCTTACGATATATCTGATATGATAT	420
Db	361	TAATTTAAAGCCCTAACTTATGATTTCCCTTTTCTTACGATATATCTGATATGATAT	420
QY	421	GCACCCATTTGTTGTCATTAATCTCCACTCTATACATCAGTATCTCAAAGTCGAATTA	480
Db	421	GCACCCATTTGTTGTCATTAATCTCCACTCTATACATCAGTATCTCAAAGTCGAATTA	480
QY	481	CAATATCCATAGAAGTGGTATATCTGAAAAAATAAATAAAGTGGTATCTGGTATA	540
Db	481	CAATATCCATAGAAGTGGTATATCTGAAAAAATAAATAAAGTGGTATCTGGTATA	540
QY	541	TACAAATACACCGTCTCGAATTCGCTCAACAAATTTAGAGAGAAATGGAAGTGTCTTT	600
Db	541	TACAAATACACCGTCTCGAATTCGCTCAACAAATTTAGAGAGAAATGGAAGTGTCTTT	600
QY	601	TGGTTTTATTTATTTTAAATACATCTATATTTTAAACACTTCGATCTCGCTTA	660
Db	601	TGGTTTTATTTATTTTAAATACATCTATATTTTAAACACTTCGATCTCGCTTA	660
QY	661	AAATTCGGAATGTCCTTAAATCTCTAAATCATAAATCGTAAAGAAATTCGTGGAAGCCA	720
Db	661	AAATTCGGAATGTCCTTAAATCTCTAAATCATAAATCGTAAAGAAATTCGTGGAAGCCA	720
QY	721	CAGGGACATGCATAGGCGAGTATGACCTTTAAACCAATCAAAATATATTAATAGAA	780
Db	721	CAGGGACATGCATAGGCGAGTATGACCTTTAAACCAATCAAAATATATTAATAGAA	780
QY	781	AGGAATCTTCTTAAAGAACAAATTTAATAAGTGGATATAAAGATTAAGAGTAGGCA	840
Db	781	AGGAATCTTCTTAAAGAACAAATTTAATAAGTGGATATAAAGATTAAGAGTAGGCA	840
QY	841	GAAAGAAAACGTATGGCGCGACTCGTAAACAAAGGAGCGTCCCGACCATCGCGAGCGCG	900
Db			

Db	841	GAAGAAAAAGTATGGCCGCGACTCGTAACAAGGACGTCCCGACCACTCGGAGACGGCG	900
QY	901	AGACGCTGACTGATTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	960
Db	901	AGACGCTGACTGATTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	960
QY	961	GTCAAAACCATATCCAATTTGTTCTGCTATATATATAAATAAAGATCCCTCTTTGTGC	1020
Db	961	GTCAAAACCATATCCAATTTGTTCTGCTATATATATAAATAAAGATCCCTCTTTGTGC	1020
QY	1021	TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTTAGTTCTTAAATATATATATCTTA	1080
Db	1021	TTTGTCTTTATTCGTGATATATAATCTAACTTAAATTTAGTTCTTAAATATATATATCTTA	1080
QY	1081	CCTATGTTTCTACTGACCTCAGTCCCTAGTATATGGACATATGGAATAAATGACGC	1140
Db	1081	CCTATGTTTCTACTGACCTCAGTCCCTAGTATATGGACATATGGAATAAATGACGC	1140
QY	1141	CCAAAATTTGAAGAGTTCTCTTCCGCAACTAACTCTTATCTTACTCATTTGAGCTATGT	1200
Db	1141	CCAAAATTTGAAGAGTTCTCTTCCGCAACTAACTCTTATCTTACTCATTTGAGCTATGT	1200
QY	1201	TAAATATCAATGTTGGCAGTCTCTCGTATTTAAATATGCCAGTTGCACCTAGATAAAAAAC	1260
Db	1201	TAAATATCAATGTTGGCAGTCTCTCGTATTTAAATATGCCAGTTGCACCTAGATAAAAAAC	1260
QY	1261	ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTTGGATTTAGCTGAT	1320
Db	1261	ATGATAGACATTTAGTTTAAAACTTGAAATGTTATTTGAACTCTTTTGGATTTAGCTGAT	1320
QY	1321	GTGTATGATTAATTTTGAAGATATTTATATTTGAAGATGTTTATATATATATATATAGAT	1380
Db	1321	GTGTATGATTAATTTTGAAGATATTTATATTTGAAGATGTTTATATATATATATATAGAT	1380
QY	1381	TTATATAGCAGAAAAATATTGATGTTAGTGTCTTTTGTCTTTTGTAGTACTCTTTTGTTC	1440
Db	1381	TTATATAGCAGAAAAATATTGATGTTAGTGTCTTTTGTCTTTTGTAGTACTCTTTTGTTC	1440
QY	1441	GTAGTCTTTTCTCTCATCTCTATGAAAGAAAAATCCAAAATAGTTTAGGAAATTTTTC	1500
Db	1441	GTAGTCTTTTCTCTCATCTCTATGAAAGAAAAATCCAAAATAGTTTAGGAAATTTTTC	1500
QY	1501	TGTAATTCATAGTCTTTTTCGTAACCAAGTCTCTATGTAAGTATCGTCATCATATTCCTC	1560
Db	1501	TGTAATTCATAGTCTTTTTCGTAACCAAGTCTCTATGTAAGTATCGTCATCATATTCCTC	1560
QY	1561	TTTGCAACCAACAAAAAGATCGTTTTTGTAAAAATTTAGTAGGCACTTAAAGTCGTCA	1620
Db	1561	TTTGCAACCAACAAAAAGATCGTTTTTGTAAAAATTTAGTAGGCACTTAAAGTCGTCA	1620
QY	1621	GTGTCTCTGCGAAATCTAGCGTTCTGTATCCACAAATAAGTTGTTGATTCGAGCTTC	1680
Db	1621	GTGTCTCTGCGAAATCTAGCGTTCTGTATCCACAAATAAGTTGTTGATTCGAGCTTC	1680
QY	1681	CAAGATTAATATCTTTTATAGTGGTCTATGAAGATTTCTAACTTCGTATACGAGTGTAT	1740
Db	1681	CAAGATTAATATCTTTTATAGTGGTCTATGAAGATTTCTAACTTCGTATACGAGTGTAT	1740
QY	1741	CCATATAATTTCTAACATATACGTCTTTTGGTAGGCTCTGCGCTTTTGGAGCCAC	1800
Db	1741	CCATATAATTTCTAACATATACGTCTTTTGGTAGGCTCTGCGCTTTTGGAGCCAC	1800
QY	1801	CCCTTGTCTTAAATGTTTGTGCACTTACATCCATATCCATATAGTTTACGAGTCCAGT	1860
Db	1801	CCCTTGTCTTAAATGTTTGTGCACTTACATCCATATCCATATAGTTTACGAGTCCAGT	1860
QY	1861	TGCACCAAAATGCTCCAAATATAAATTTAAATTTGGCCACAAAACATTTTACAAACAA	1920
Db	1861	TGCACCAAAATGCTCCAAATATAAATTTAAATTTGGCCACAAAACATTTTACAAACAA	1920
QY	1921	ATTCAACAAACATGCTATCGTTTCAAAATTTATTTTCAATGCGGTTATTTGTTTCA	1980
Db	1921	ATTCAACAAACATGCTATCGTTTCAAAATTTATTTTCAATGCGGTTATTTGTTTCA	1980

QY 1981 AAATAATCTGTTTAACTCACTGACGAATTTTTTAAATTTTCAAAGAAGAACATTTTTTGAT 2040
DB |||||
1981 AAATAATCTGTTTAACTCACTGACGAATTTTTTAAATTTTCAAAGAAGAACATTTTTTGAT 2040
QY 2041 ATAATAACATTTTATGAAACACCGGTTAAGCTCGATGATTTTGAGTTTGTGTC 2100
DB |||||
2041 ATAATAACATTTTATGAAACACCGGTTAAGCTCGATGATTTTGAGTTTGTGTC 2100
QY 2101 GTTTTGTGAATCATTAAAGCCTACATTTTGATCCTCATTAATCTTAAATATAGGAATC 2160
DB |||||
2101 GTTTTGTGAATCATTAAAGCCTACATTTTGATCCTCATTAATCTTAAATATAGGAATC 2160
QY 2161 AAACATGATGATTAAGTTTCAACAAAGAGCGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
DB |||||
2161 AAACATGATGATTAAGTTTCAACAAAGAGCGTCTCTTATGGCTATTAAGAGTCAGACGCAAG 2220
QY 2221 GATGACCGGGTCATTAAAGCGTCTTATATTC AACCATTAATCTCAAGCAACAGGC 2280
DB |||||
2221 GATGACCGGGTCATTAAAGCGTCTTATATTC AACCATTAATCTCAAGCAACAGGC 2280
QY 2281 TCAGATTAATTTGTTTAAATACGATAATGTTATTTGATTAAGTAGTCTTCAGCAACAGGC 2340
DB |||||
2281 TCAGATTAATTTGTTTAAATACGATAATGTTATTTGATTAAGTAGTCTTCAGCAACAGGC 2340
QY 2341 AAAGATTAATTTGATTAATTC AAAGATTTGTGGCTTCCAAAGCATAGGGAATGGCT 2400
DB |||||
2341 AAAGATTAATTTGATTAATTC AAAGATTTGTGGCTTCCAAAGCATAGGGAATGGCT 2400
QY 2401 CCACATGATGGGAATATATTAATTAATTAATTC AACCTCCATTTGCCACAGTCGTTGGA 2460
DB |||||
2401 CCACATGATGGGAATATATTAATTAATTAATTC AACCTCCATTTGCCACAGTCGTTGGA 2460
QY 2461 GTTTATTTTTTTTTTCCAAAGTGGGAAGATCAATTAATAATTTGCGGAATTTCTAATCTCTA 2520
DB |||||
2461 GTTTATTTTTTTTTTCCAAAGTGGGAAGATCAATTAATAATTTGCGGAATTTCTAATCTCTA 2520
QY 2521 CTTGTTGAAACAAACCCAAAGCAAAATTAAGTTTGAAGAATGTAACGGAATCTACTATAGA 2580
DB |||||
2521 CTTGTTGAAACAAACCCAAAGCAAAATTAAGTTTGAAGAATGTAACGGAATCTACTATAGA 2580
QY 2581 ATTATCTAAACATATCATATAGCTTTGACTTTTAAATTAATAATTAATAATTTGCGGCA 2640
DB |||||
2581 ATTATCTAAACATATCATATAGCTTTGACTTTTAAATTAATAATTAATAATTTGCGGCA 2640
QY 2641 TTAGTTAGATATTTTACCAGGGGAATTAAGCAGCACTGTTCAATGCACTCTCTCTTAATTT 2700
DB |||||
2641 TTAGTTAGATATTTTACCAGGGGAATTAAGCAGCACTGTTCAATGCACTCTCTCTTAATTT 2700
QY 2701 ACTATCTTTGAAAGAACTTATPATGCTCACTGTAATGCCAGTTGCCACTTAATATAAACA 2760
DB |||||
2701 ACTATCTTTGAAAGAACTTATPATGCTCACTGTAATGCCAGTTGCCACTTAATATAAACA 2760
QY 2761 ACATTTCACTTTGTCACATCGCTGTAATGAATGAATTTGGAACGACCTCTTAACACTAAT 2820
DB |||||
2761 ACATTTCACTTTGTCACATCGCTGTAATGAATGAATTTGGAACGACCTCTTAACACTAAT 2820
QY 2821 AGGGTTAATTAACCACTAAATTTTTTCCAAAGTTGTCATTTTGTCTTAATGTCGAGCGTAAT 2880
DB |||||
2821 AGGGTTAATTAACCACTAAATTTTTTCCAAAGTTGTCATTTTGTCTTAATGTCGAGCGTAAT 2880
QY 2881 APTCAATACGGGTCTAACCTAAGAGTTGGTTCGATCACAAATTTTGAAGATACCT 2940
DB |||||
2881 APTCAATACGGGTCTAACCTAAGAGTTGGTTCGATCACAAATTTTGAAGATACCT 2940
QY 2941 TTCATTAATAAATTTTGTGTTATATTTTCAACCGTGTAAAGTAAACCTTAATTAGGAGCT 3000
DB |||||
2941 TTCATTAATAAATTTTGTGTTATATTTTCAACCGTGTAAAGTAAACCTTAATTAGGAGCT 3000
QY 3001 ATTTTCTATTTGATGTAATTTGAAATTTGTAATTTCTATATAAATTAAGTATGGAAGGAATG 3060
DB |||||
3001 ATTTTCTATTTGATGTAATTTGAAATTTGTAATTTCTATATAAATTAAGTATGGAAGGAATG 3060

QY 3061 TAAATTAATATAGAACACAGATACAAAAAGATGTCCCGTGCTTTAAACACGCTGAGTCATTG 3120
DB |||||
3061 TAAATTAATATAGAACACAGATACAAAAAGATGTCCCGTGCTTTAAACACGCTGAGTCATTG 3120
QY 3121 TCGTACCCCTTTTGGCAACTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTAAAA 3180
DB |||||
3121 TCGTACCCCTTTTGGCAACTTTTCAAGTTTCTTTCGTGAAAATGACTACACTTTTAAAA 3180
QY 3181 TAAATTGACAGATGATTTGTCATGCATATAATATTTCCGCAAAATGCCAAATCTACCCCTT 3240
DB |||||
3181 TAAATTGACAGATGATTTGTCATGCATATAATATTTCCGCAAAATGCCAAATCTACCCCTT 3240
QY 3241 AACCAAAATAATGGGTAAATGGATATAAAATAGTTAAACATAAAATTTGCAAAATTTTG 3300
DB |||||
3241 AACCAAAATAATGGGTAAATGGATATAAAATAGTTAAACATAAAATTTGCAAAATTTTG 3300
QY 3301 AAAAATAGAGTAATAATTTGATTTTTTTTAAAGTTTGAATTTGAAGTGA AAAAATATATA 3360
DB |||||
3301 AAAAATAGAGTAATAATTTGATTTTTTTTAAAGTTTGAATTTGAAGTGA AAAAATATATA 3360
QY 3361 ATAAAAATAATAACCTGTAGTTTATATATATATAGTTAGTAGTCAAGTTTGAGTAACT 3420
DB |||||
3361 ATAAAAATAATAACCTGTAGTTTATATATATATAGTTAGTAGTCAAGTTTGAGTAACT 3420
QY 3421 TGAAGTCTTGGAATTAATTTTATATATGTTTTTCTCACAGATATATATTTTCGTTCTATCCC 3480
DB |||||
3421 TGAAGTCTTGGAATTAATTTTATATATGTTTTTCTCACAGATATATATTTTCGTTCTATCCC 3480
QY 3481 AAGAAATGGGATAATTTCTCTATATTCGAGGCTCTCTCTTTAAGAAATCGGTGATATAAT 3540
DB |||||
3481 AAGAAATGGGATAATTTCTCTATATTCGAGGCTCTCTCTTTAAGAAATCGGTGATATAAT 3540
QY 3541 CTTAAGCCCTTACTTTGACACAAAGGCTCTATATAAAAGCCCAATAATAATTTTCTTTTC 3600
DB |||||
3541 CTTAAGCCCTTACTTTGACACAAAGGCTCTATATAAAAGCCCAATAATAATTTTCTTTTC 3600
QY 3601 AAAGCCCAACACGTCAGAGGAGAAAGTGGCTTTCGTTTGAATTTGAAAACGTGGCG 3660
DB |||||
3601 AAAGCCCAACACGTCAGAGGAGAAAGTGGCTTTCGTTTGAATTTGAAAACGTGGCG 3660
QY 3661 GGCGGTTGGAATCTTAACCCCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCT 3720
DB |||||
3661 GGCGGTTGGAATCTTAACCCCTTCACTCACTCACTCACTCACTCACTCACTCACTCTCT 3720
QY 3721 GAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
DB |||||
3721 GAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
QY 3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGCTAACACTTCAGCGCCACATTT 3840
DB |||||
3781 GCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGCTAACACTTCAGCGCCACATTT 3840
QY 3841 GGGTTTACAGATAGGGTTTCAATAGATCGAATCGAATTCATCTCTCTGTTAGTATTAGATT 3900
DB |||||
3841 GGGTTTACAGATAGGGTTTCAATAGATCGAATCGAATTCATCTCTCTGTTAGTATTAGATT 3900
QY 3901 CATCAAAATCAAAATCTATCGTCACCTAAAAATCTCTTTCCGTTAGTATTATTCGGG 3960
DB |||||
3901 CATCAAAATCAAAATCTATCGTCACCTAAAAATCTCTCTTTCCGTTAGTATTATTCGGG 3960
QY 3961 GTTGTTCAAAAGTCGCTTATGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
DB |||||
3961 GTTGTTCAAAAGTCGCTTATGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
QY 4021 CTATCTAGAGAAAGAACAAATCTTTGAAATCGAAAGGTCCTAATGTAAGTTCAATGTCT 4080
DB |||||
4021 CTATCTAGAGAAAGAACAAATCTTTGAAATCGAAAGGTCCTAATGTAAGTTCAATGTCT 4080
QY 4081 ACATTATGAGATTCGCAATGATTTATAGTCAAAAGATTTCTCTAAAGCGTTTACTTTTGT 4140
DB |||||
4081 ACATTATGAGATTCGCAATGATTTATAGTCAAAAGATTTCTCTAAAGCGTTTACTTTTGT 4140
QY 4141 GGAGCATTTCTTGTAGTTCTTATGCAAAATAAAGTTCTAGGAATGATAATTTCTTAAGGAAG 4200

Db 3421 TGAAGTCTGAATACCTTATATATGTTTTTCTCAGAGATTATATATTTCTGCTATPCCC 3480
QY 3481 AAGAATTTGGGATAATATTTCTCTATATTTGAGGGCTCTCTCTTAAGAAGTCGGGTATAAT 3540
Db 3481 AAGAATTTGGGATAATATTTCTCTATATTTGAGGGCTCTCTCTTAAGAAGTCGGGTATAAT 3540
QY 3541 CTTAAGCCCTTACTTGACAAAGCCCTCTATTTAAAAGCCCAATATAATTTTCTTTTC 3600
Db 3541 CTTAAGCCCTTACTTGACAAAGCCCTCTATTTAAAAGCCCAATATAATTTTCTTTTC 3600
QY 3601 AAAGCCCAACACGCTCAAGAGGAGAAAGAGTGGTTTGGCTTTGGATTGAAAACGTTGGCG 3660
Db 3601 AAAGCCCAACACGCTCAAGAGGAGAAAGAGTGGTTTGGCTTTGGATTGAAAACGTTGGCG 3660
QY 3661 GCGGTTTGGAAATCTTAACCCCTTCACATTCATCTTCAACCGTCAAACTCT 3720
Db 3661 GCGGTTTGGAAATCTTAACCCCTTCACATTCATCTTCAACCGTCAAACTCT 3720
QY 3721 GAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGCTTTGACATCA 3780
Db 3721 GAACTGAGCTTCGACGATGTTGGGTTTACGAAGATCTGCGACGCTTTGACATCA 3780
QY 3781 GCCAGTCTCTGCTTCTGTAATGTTACGGTTAGTATCTCTTAAACATTCAGCGCCACATTT 3840
Db 3781 GCCAGTCTCTGCTTCTGTAATGTTACGGTTAGTATCTCTTAAACATTCAGCGCCACATTT 3840
QY 3841 GGGTTTACGATTAGGTTTCAATAGATCGAATCGAATTCATCTCTTGGTATTAGATTT 3900
Db 3841 GGGTTTACGATTAGGTTTCAATAGATCGAATCGAATTCATCTCTTGGTATTAGATTT 3900
QY 3901 CATCAATTCAAATCTATGTCACCTAAATAATCTCTCTTCGGTATCTAGTATTCGGG 3960
Db 3901 CATCAATTCAAATCTATGTCACCTAAATAATCTCTCTTCGGTATCTAGTATTCGGG 3960
QY 3961 GTTGTTTCAAAGTCGCTTATGCTCTGTGATTTGGCTTCTTACCCATTTTCAACAGTG 4020
Db 3961 GTTGTTTCAAAGTCGCTTATGCTCTGTGATTTGGCTTCTTACCCATTTTCAACAGTG 4020
QY 4021 CTATGTAGAAAGAAACAAATCTTTGAAATCGAAAGGCTTAATGTATAGTTCAAATGCT 4080
Db 4021 CTATGTAGAAAGAAACAAATCTTTGAAATCGAAAGGCTTAATGTATAGTTCAAATGCT 4080
QY 4081 ACATTATGAGATGCGCATGATATATAGTCAACAGATTCCTTAAAGCGTTACTTTGT 4140
Db 4081 ACATTATGAGATGCGCATGATATATAGTCAACAGATTCCTTAAAGCGTTACTTTGT 4140
QY 4141 GGAGCAATCTTTGTTAGTTCTATGCAAAATAAAGTTCTAGGAATGATAAATCTTAAAGAG 4200
Db 4141 GGAGCAATCTTTGTTAGTTCTATGCAAAATAAAGTTCTAGGAATGATAAATCTTAAAGAG 4200
QY 4201 CATCTCAATGTTGGCTAGTTCTTGCTCAGGTTTAAACAAATGTTTGCATTTGCTTTT 4260
Db 4201 CATCTCAATGTTGGCTAGTTCTTGCTCAGGTTTAAACAAATGTTTGCATTTGCTTTT 4260
QY 4261 AGTTAGAAATGTTGACTTGCTGAGTTTGTGCTTAACTGCTCTGTTGAAGCAAAAGTTT 4320
Db 4261 AGTTAGAAATGTTGACTTGCTGAGTTTGTGCTTAACTGCTCTGTTGAAGCAAAAGTTT 4320
QY 4321 GATCAAAACCATTAACCTTTATTTGATTTCTCTTTTGAATATATACATCAATTTATGATTT 4380
Db 4321 GATCAAAACCATTAACCTTTATTTGATTTCTCTTTTGAATATATACATCAATTTATGATTT 4380
QY 4381 TCTTTCTCTACAGTTTTCATGGGTTACAGTCCA 4413
Db 4381 TCTTTCTCTACAGTTTTCATGGGTTACAGTCCA 4413

RESULT 3

AC012193/c 82454 bp DNA linear PLN 19-JAN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence,
DEFINITION complete sequence.

AC012193
VERSION
KEYWORDS
SOURCE
ORGANISMREFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

FEATURES

source

1. 82454
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T32E8"
complement(1..3320)
/gene="T32E8.1"
/note="similar to ribonuclease II RNB family protein
GB:CA16367 [Schizosaccharomyces pombe]; Pfam HMM hit:
VacB and RNase II family 3'-5' exoribonucleases"

gene

mRNA

CDS

complement(join<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..3320))
/gene="T32E8.1"
complement(join<1..69,154..2280,2381..2449,2524..2577,
2676..3055,3215..3320))
/gene="T32E8.1"
/codon_start=1
/product="putative 3'-5' exoribonuclease, 3' partial;
3320-1"
/protein_id="AAG51632.1"

AC012193 GI:12323286
HTG.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Eukaryota; Magnoliophyta; eudicotyledons; core eudicots;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eudicots II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 82454)
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Renning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence
Unpublished
2 (bases 1 to 82454)
Lin,X. and Kaul,S.
Direct Submission
Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 82454)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280880.
Address all correspondence to:at@tigr.org

BAC clone T32E8 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Location/Qualifiers

	misc_feature	/db_xref="GI:12323300" /translation="MKSASSEQSVIERIENGHKRRNPQKONRRSKOSSVPIEDAHVE ESLDGRSSKADSTSSSKQKQPNPTDELEAMASNAFNMPPMRAESGYPRRSAS PLSSPVSQKLLSKCSPDPRACQSPGMGELPQQLEGSSQRKI FSSHWSLDAVTEA LEKGEAFKALFRVNAHRNAYCKIDGVPTDILINGVCSRAVEGDTVIKLDPLSL WPKMGVTSAAKPEGTNSPPEKDKKARKQNGIDVVEGFEDGFSKNKSSVLKGGAK NGVTPSPPLSDSCIGFCQKGNCSADVCLCGILSFPKHKPTGQVAVVSKSLVRD SIVGLVDKGMHIIKYESDKPKCKPLSLSDDEYVQLMPADPRFKLLVPPHVLPGSIR ARLENDLNEALELAAQIVDMGEGSPFVAQI THLFGSGELPQALNIQNSVCD SDSPGSLTSLPRVPEVPEEVEORRDLRDLCVLTIDPSTATDLDAISVOSLPGGF PRVGHADVSFVLPETALDTEARFSTSVYLMQKISMPLPILLSSEVQSLSPGAOR LAFSILMDLNRGVIDRWIGRTIIRSCKLSYDHAQIIDGKSDVAENGVPAUHGSR KWDVTSVQKLSAISTTLQKRFRRNGALQLENSKVPFLKEFGVCPVYFVCSRKGSN FLVEFMLLANMTAAEYISQAYRASLLRRHPENRKLKEFEFCGSKHGMDLDSIR GLODSLEKTIQNLKDSVFDVIDLNNVAIKPMOLASVFCNGLKDSVAEWGHYALAVP LYTHFTSPLRPDIDVHRALAALEAEELYSKOKOTAIDEGRSCFTGIHFNKDAES IEGKEALSVAALKHGVSTELISDVAAYCHERKLAARKVRDACDKLYTFWVLKQKEIF PCEARVNNLGRFMIVVISKL"	
		1. .189 /note="overlap with BAC clone T5M16 [AC010704:103165..103353]." complement (8589..10716) /gene="T32E8.2" /note="similar to AUX1 (regulator of root gravitropism, putative permease) GB:CAA67308 [Arabidopsis thaliana] complement [join(<8589..8783,9000..9465,9555..9652, 9753..9936,10020..10132,10236..10421,10504..>10716)] /gene="T32E8.2" complement [join(8589..8783,9000..9465,9555..9652, 9753..9936,10020..10132,10236..10421,10504..10674)] /gene="T32E8.2" /codon_start=1 /product="putative AUX1-like permease; 10674-8589" /protein_id="AAG51630.1" /db_xref="GI:12323298"	
	gene	/translation="MAAEKRTVAVAGNYLEMEREBENISGNKKSSTTKYLSNFPWHGG SVYDAFSCASNOAVQLLTLFYSFQSGMMGSLPQLFGLMGSTAYLISVLVVEY RTRERKEKDFRNHVIOMFEVLIDGLGKHWRNLGLI FNCTFLFGSVIQLIACASNIY YINDLKRTVYTIYFGACATVTPGFHNRYIWSFLGLAMTVTSWYLTATSLHGG AEDVKHSGPTMWLYFTGATNLYTFGHAVTVEIMHAMWKPQKFAIYLLATLYVLT LTPSASAVYAFGDKLLTHSNALSILPKTGFRDTAVILMLIHQFIFFGEASTPLFYV WEKLGIVHETKSMFKRAVARLPVVPIWLAIIFFPFGPINSVAGSLVSTFYIIIPA LAHMLTFAPASRENAVERPRVVGWNGTYCINIFVVVVFVVGFGGWSMVNFV RQIDTFGLFTKCYQCPHPK" 12104..13574 /gene="T32E8.3" /note="similar to thaumatin-like protein GB:AAB71214 [Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family" join(<12104..12179,12505..12628,12704..>13574) /gene="T32E8.3" join(12104..12179,12505..12628,12704..13574) /gene="T32E8.3" /codon_start=1 /product="thaumatin-like protein; 12104-13574" /protein_id="AAG51631.1" /db_xref="GI:12323299"	
		/translation="MDRTIEIEFDINGPYDSLIRPNISVASLTSRSENNITFPDKIFS SKDASFRFVKDLVTLSSHITRVMYITIRTPVNYSTGPKWSEARVFTI INSCDQTIW PAITPGNFNGGVELKPGQSIIVHAPVWGSRINGRTGCKEDTGTGTCGTCGSGST LKCSAGPAPSLAEFTLAALDFDYDLSLDGFLPMSVTPMNGKNGSVAGCADLRP HCPQELAVKNGKVISCRSADVFDRDEYCCRGVYGVNVCQPTYYSKIFKQACPTAY SYAYDDPTS IMTASDADYIVSFCSSRFVLFLETTSHDYFNYSNQVSFIIDHDFYIY MIYFTHPLLIISYAKVKVIQMQK" 14054..15303 /gene="T32E8.4" /note="similar to hypothetical protein GB:P34661 [Caenorhabditis elegans]" join(14054..14174,14251..14348,14779..14811,15170..15303) /gene="T32E8.4" join(14107..14174,14251..14348,14779..14811,15170..15252) /gene="T32E8.4" /codon_start=1 /product="unknown protein; 14107-15252"	
	mRNA	/protein_id="AAG51633.1" /db_xref="GI:12323301" /translation="MATGKQVSPKVTLTSDPKLPFKVSPVEGAPFTAVLKFAAEFEK VPQTSAILIINDGIGINPQSQAGNVFLKHGSELRLIPDRVGAVFVMDP" 18583..21388 /gene="T32E8.5" /note="Pfam HMM hit: Eukaryotic protein kinase domain, C-terminal region is identical to putative protein kinase GB:CAA72680 [Arabidopsis thaliana]" join(<18583..18916,19056..19948,20068..20268,20382..20543, 20662..21388) /gene="T32E8.5" join(18916,19056..19948,20068..20268,20382..20543, 20662..21388) /gene="T32E8.5" /codon_start=1 /product="putative protein kinase; 18583-21234" /protein_id="AAG51619.1" /db_xref="GI:12323287"	
		/translation="MQTTSIGPRRSVAPQASRNTRLVTAEGRSQDVVTLQSGLAA NTLTQDTNNLAITSVAGESASITQPTVSEHFNPSDRQMDFGSAVTSLSNLDVQRK SOSLIGTSQDMEWDATNOAEASHLDACIGSKHONLPVSDSEVSLKSVKSSSLAKIQ GOLGEFENFLNQPRTCSRCAVSSSWATTLIHSSAPMLNATTHVSRVSVEADSNAPH AVSQGNLPSCCFSSKVNILHPNKDATASEMPASTNDPEVRVKEITDSKQQIITGL EAPVSSYISGDSQANARLPEELHITSVSSQPKSDKHEKVASSKGSAPRKRNYDDPL FFKVGKLYQRLGKIGSGSEVHKVISSDCTIYALKIKLKGSDATAYAGFCQEIGY LKKLKGKNTIQLIDYEVTDLQVINGTMSNKDGRVKEDGFIYMWLEYGEIDLAK MLSKWREIEGSDRTIDENLRFYWOQILOAVNTIHERIVHSDLKPANFLVLRGFLK LIDFGIAKAINSDTNIQRDSQVCTLSYMSPEAFMCNDESDENGNTIKGSPSDIWSLG CILYQMVYGRTPADYKTFWAKFKVITDPRHEIYNGLSNFWLIDLKKKCLAWBRNQ WRPELLOHPFLAPPPIPEPOVKTKILFLIAESGSDDDDKANSMISOLEQLLSNPAP LPRNDVLDSDQNQLLSRVSSELQIQDLRQLQGLEGN" 21802..22921 /gene="T32E8.6" /note="similar to putative SWH1 protein GB:BAC20736 [Arabidopsis thaliana]; Pfam HMM hit: PH (pleckstrin homology) domain" join(<21802..22203,22526..>22921) /gene="T32E8.6" join(21802..22203,22526..22921) /gene="T32E8.6" /codon_start=1 /product="hypothetical protein; 21802-22921" /protein_id="AAG51621.1" /db_xref="GI:12323289"	
	mRNA	/translation="MSSPLAAKSNLTTRASLNI PHSTRNNHDLKRYLSAGPRTN EDRPASNGVAGILYKWNVYQCKRWFVLDGVLVYRIHGPKTISLVEMDRRSK LIGGESURFICRHSKGVDSHPGLQIHLKVSIGQSISDGKRFVFTGTGKSLHLR AATSEDRASMI EALKAVKETFPRMSNEELMASTTNVSVSTDKLKRLMEEBVDETIK DCEDIMKNFIALHDEYMSLKRYQYHLIDSLKNVSNSPIRPSNQSFS" 27895..31218	
		Query Match 91.8%; Score 4052; DB 8; Length 82454; Best Local Similarity 99.9%; Pred. No. 0; Matches 4412; Conservative 0; Mismatches 0; Indels 3; Gaps 3; QY 1 AAGATCCACAGTGAATAATAAAGAACGGAATTCGGTGATATTCGCACTATAATGAA 60 Db 42585 AAGATCCACAGTGAATAATAAAGAACGGAATTCGGTGATATTCGCACTATAATGAA 42526	
	QY	61 ATTGAATCTCTGATTCATCGCTTTTGTATCAAGATCGAATCTCTAAAAACATATCTCTA 120	
		Db 42525 ATTGAATCTCTGATTCATCGCTTTTGTATCAAGATCGAATCTCTAAAAACATATCTCTA 42466	
	QY	121 TAAATAATCTCGAGTCGAATTCGTATATATTCGACATCTGAAAGATTAAGTGGG 180	
		Db 42465 TAAATAATCTCGAGTCGAATTCGTATATATTCGACATCTGAAAGATTAAGTGGG 42406	
	QY	181 ATATATAACAAACCGCGGAAAGTACAACTTCTACTCTTTTGGCATGGAAACCATGTTTT 240	
		Db 42405 ATATATAACAAACCGCGGAAAGTACAACTTCTACTCTTTTGGCATGGAAACCATGTTTT 42346	
	QY	241 TAGGATTACTTTTGTAATCTCTGAATCTTTCATTCTTGTGAATGATATTTACATTTTAT 300	

QY	2461	GTATTATTTTTTTTCCAAAGTGGAGAAATCAATTATAATTGTCGGAATTTCTAAATCCTA	2520
Db	40125	TTTTTTTTTTTTTTCCAAAGTGGAGAAATCAATTATAATTGTCGGAATTTCTAAATCCTA	40066
		GTTTATTTTTTTTTTTCCAAAGTGGAGAAATCAATTATAATTGTCGGAATTTCTAAATCCTA	40066
QY	2521	CTTGTTGGAAAAAACAACCCAGCAAAAATTAGTTTGAAGATGTACGGAATCTCTACTATAGA	2580
Db	40065	CTTGTTGGAAAAAACAACCCAGCAAAAATTAGTTTGAAGATGTACGGAATCTCTACTATAGA	40006
QY	2581	ATTATACTAAACATATCAATATAGCTTTTGACTTTAAATTTAAACAAATATTATGTGGGCAA	2640
Db	40005	ATTATACTAAACATATCAATATAGCTTTTGACTTTAAATTTAAACAAATATTATGTGGGCAA	39946
QY	2641	TTAGTTTAGATATTTTTTACCAGGGGAAATAAGCAGCACTGTTCAATGCACCTCTCTTCTTAATT	2700
Db	39945	TTAGTTTAGATATTTTTTACCAGGGGAAATAAGCAGCACTGTTCAATGCACCTCTCTTCTTAATT	39887
QY	2701	ACTATACCTTTGGAAGAACCTATATATGTCACCTGTATTTGCCAGTGGCCACTAATATAAACA	2760
Db	39886	ACTATACCTTTGGAAGAACCTATATATGTCACCTGTATTTGCCAGTGGCCACTAATATAAACA	39827
QY	2761	ACATTTCACTTGTTGATACGCTGTAAATGAAGTTTGGAAACACCCCTCTTAACAACTAAT	2820
Db	39826	ACATTTCACTTGTTGATACGCTGTAAATGAAGTTTGGAAACACCCCTCTTAACAACTAAT	39767
QY	2821	AGGGTTAAATTAACCACTAAATTTTTTCCAAGTTGTCAATTTTGTCTTAATGTGAGACGTAAT	2880
Db	39766	AGGGTTAAATTAACCACTAAATTTTTTCCAAGTTGTCAATTTTGTCTTAATGTGAGACGTAAT	39707
QY	2881	ATCTAATACGTCGGTCTTAACCTTAAGAGTTGGTTCCGATCACAAAAATTTTGTGAGAGTACCT	2940
Db	39706	ATCTAATACGTCGGTCTTAACCTTAAGAGTTGGTTCCGATCACAAAAATTTTGTGAGAGTACCT	39647
QY	2941	TTCAATTAACAAAAATTTGTGTATTAATTTCCACGCTGTTAAGTAAAAACTTAATTAGAGCT	3000
Db	39646	TTCAATTAACAAAAATTTGTGTATTAATTTCCACGCTGTTAAGTAAAAACTTAATTAGAGCT	39587
QY	3001	ATTTTCTATTTGATGTGAATTTGAAAAATGCTTCATAAAAAATAGTATGGAAGAGGAATG	3060
Db	39586	ATTTTCTATTTGATGTGAATTTGAAAAATGCTTCATAAAAAATAGTATGGAAGAGGAATG	39527
QY	3061	TAATTAATATAGACACACAGATACAAAAGATGTCCCGTCTTTAAACACGCTCGAGTCATTG	3120
Db	39526	TAATTAATATAGACACACAGATACAAAAGATGTCCCGTCTTTAAACACGCTCGAGTCATTG	39467
QY	3121	TCGTACCCCTTTTGCCAACTTTTTCAGTTTCTTTTCGTGAAAATGACTACACTTTTAAAA	3180
Db	39466	TCGTACCCCTTTTGCCAACTTTTTCAGTTTCTTTTCGTGAAAATGACTACACTTTTAAAA	39407
QY	3181	TAAATTTGACAGATGATTTGATGCAATATAATATTCGCCAAAATGCCAAAATTCACCCCTT	3240
Db	39406	TAAATTTGACAGATGATTTGATGCAATATAATATTCGCCAAAATGCCAAAATTCACCCCTT	39347
QY	3241	AACCAATAATG-GGTAATGGATATAAAATAGTTAAACATAAAACAAAAATTTTGGAAATTTT	3299
Db	39346	AACCAATAATGTTGGTATATGGAATATAAAATAGTTAAACATAAAACAAAAATTTTGGAAATTTT	39287
QY	3300	GAAAAAATAGAGATAAATGATTTTTTTTTTAAAAAGTTTGAATGGAATGAAAAATAATAT	3359
Db	39286	GAAAAAATAGAGATAAATGATTTTTTTTTTAAAAAGTTTGAATGGAATGAAAAATAATAT	39227
QY	3360	-AATAAAAAATAAACCCTGAGTTTGATATATATATAGTTAGATAGCTCAAGTTTGAGTAA	3418
Db	39226	AAATAAAAAATAAACCCTGAGTTTGATATATATATAGTTAGATAGCTCAAGTTTGAGTAA	39167
QY	3419	CTTGAAGTCTTGAAATACCTTTATATGTTTTTCTCACAGATTAATTAATTTTGGTCTATC	3478
Db	39166	CTTGAAGTCTTGAAATACCTTTATATGTTTTTCTCACAGATTAATTAATTTTGGTCTATC	39107
QY	3479	CCAAGAAATGGGATAATATCTCTATATTCGAGGCTCTCTCTTAAGAAAGTCGGTGTATA	3538
Db	39106	CCAAGAAATGGGATAATATCTCTATATTCGAGGCTCTCTCTTAAGAAAGTCGGTGTATA	39047

Qy	3539	ATCTTAAGCCCTTACTTGGACACAGGCGCTCTATTAAAAAGCCCAATAAATATTTTCTTT	3598
Db	39046	ATCTTAAAGCCCTTACTTGGACACAGGCGCTCTATTAAAAAGCCCAATAAATATTTTCTTT	38987
Qy	3599	TCAAAGCCCAACACGTCACAGAGGAGAAAGAGTGGCTTTGGCTTTGGATTCGAAACAGTGG	3658
Db	39886	TCAAAGCCCAACACGTCACAGAGGAGAAAGAGTGGCTTTGGCTTTGGATTCGAAACAGTGG	39927
Qy	3659	CGGCGGGTTGGAAACTGAATCTTTAAACGCTTCACATTCATCTTCCACGTCGTAATCT	3718
Db	38926	CGGCGGGTTGGAAACTGAATCTTTAAACGCTTCACATTCATCTTCCACGTCGTAATCT	38867
Qy	3719	CTGAAACTGAGCTTCGACGATCTCGGGTTACGAAGATCTCGACGACCTTGTTCGCAT	3778
Db	38866	CTGAAACTGAGCTTCGACGATCTCGGGTTACGAAGATCTCGACGACCTTGTTCGCAT	38807
Qy	3779	CAGCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAAACACTTCAGCGCCACAT	3838
Db	38806	CAGCCAGTCTCTGCTTCGTAATGTTACGGTTAGTATCTGTCTAAACACTTCAGCGCCACAT	38747
Qy	3839	TTGGGTTTACGATTAGGGTTTCAATAGATCGAATTCGATTCATCTCTCTTGGTATTAGAT	3898
Db	38746	TTGGGTTTACGATTAGGGTTTCAATAGATCGAATTCGATTCATCTCTCTTGGTATTAGAT	38687
Qy	3899	TTCAATCAAAATCAAAATCTATCGTCACCTTAAAAATCTCTCTTTGGTATGTAGTATTTCCG	3958
Db	38686	TTCAATCAAAATCAAAATCTATCGTCACCTTAAAAATCTCTCTTTGGTATGTAGTATTTCCG	38627
Qy	3959	GGGTTGTTTCAAAGTCGCTTATATGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAG	4018
Db	38626	GGGTTGTTTCAAAGTCGCTTATATGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAG	38567
Qy	4019	TGCTATGTAGAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATATAGTTCAAATGT	4078
Db	38566	TGCTATGTAGAGAAAGAACAAATCTTTGAAATCGAAAGGTCTAATGTATATAGTTCAAATGT	38507
Qy	4079	CTACATATTAGATTCGCCATGATATTATAGTCAAAACGATTTCTCTTAAAGCGTTTACTTTTT	4138
Db	38506	CTACATATTAGATTCGCCATGATATTATAGTCAAAACGATTTCTCTTAAAGCGTTTACTTTTT	38447
Qy	4139	GTGAGCATCTTTGTTAGTCTTATGCAAAATAAGTTCCTAGGAATGATAATCTTTAAGGA	4198
Db	38446	GTGAGCATCTTTGTTAGTCTTATGCAAAATAAGTTCCTAGGAATGATAATCTTTAAGGA	38387
Qy	4199	AGCATCTCAAAATTTGGCTAGTTCCTGTCTCAGGTTTAAACAAATGTTTTCGAAATTTGCTT	4258
Db	38386	AGCATCTCAAAATTTGGCTAGTTCCTGTCTCAGGTTTAAACAAATGTTTTCGAAATTTGCTT	38327
Qy	4259	TTAGTTAGAAATTTGTTGACTTGTCTAGCTTTTGTGACTCTGCTCTGTGTGAGCAAAAGT	4318
Db	38326	TTAGTTAGAAATTTGTTGACTTGTCTAGCTTTTGTGACTCTGCTCTGTGTGAGCAAAAGT	38267
Qy	4319	TTGATCAAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTTATACATCAATTTATGTAT	4378
Db	38266	TTGATCAAAACCCATTACCTTATTTGATTTCTCTCTTTAGATTTATACATCAATTTATGTAT	38207
Qy	4379	TTTCTTTGTCTACAGTTTTCATGGGTTACGAGTCCA	4413
Db	38206	TTTCTTTGTCTACAGTTTTCATGGGTTACGAGTCCA	38172

RESULT	4
AX063412	
LOCUS	
DEFINITION	Sequence 4 from Patent WO0100833.
ACCESSION	AX063412
VERSION	AX063412.1 GI:12541200
KEYWORDS	.
SOURCE	synthetic construct
ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	1 Hoffmann,B., Mollner,P. and Pelletier,G.
PATENT	linear PAT 24-JAN-2001

TITLE Promoter expressed specifically in the cells of plant roots,
recombinant vectors and host cells comprising same and transgenic
plants obtained
JOURNAL Patent: WO 0100833-A 4 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES Location/Qualifiers
source 1..4309
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Insert du vecteur pBin19"

ORIGIN

Query Match 48.8%; Score 2154; DB 6; Length 4309;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 CTGCAGTCGAATTTGTGATATATTGTAAGCAATCTGAAGAAATGAAGTGGGATATATAAAC 190
DB 25 CTGCAGTCGAATTTGTGATATATTGTAAGCAATCTGAAGAAATGAAGTGGGATATATAAAC 84

QY 191 AACGGCGAAAGTACAAGTCTACCTTTTGGCATGGAACCAATGTTTTAGGATTTAC 250
DB 85 AACGGCGAAAGTACAAGTCTACCTTTTGGCATGGAACCAATGTTTTAGGATTTAC 144

QY 251 TTGTGAATTCCTGAATCTTTTCATTTCTTGAATTGATATTTACATTTTATCAAAAAAAA 310
DB 145 TTGTGAATTCCTGAATCTTTTCATTTCTTGAATTGATATTTACATTTTATCAAAAAAAA 204

QY 311 GTACAAGTCTTACCAAGCACAGGAGTTAAACAACCTTGTGTGCAAAATGCTAAATTTAAG 370
DB 205 GTACAAGTCTTACCAAGCACAGGAGTTAAACAACCTTGTGTGCAAAATGCTAAATTTAAG 264

QY 371 CCTAATCTTAGATTTCCCTTTTCTTCAAGATATATCTGATATTTGATATGACCCCATTT 430
DB 265 CCTAATCTTAGATTTCCCTTTTCTTCAAGATATATCTGATATTTGATATGACCCCATTT 324

QY 431 GTTGTGATTAACCTTCCCTCTATACATCAGTATCTCAAGTCGATACCAATATCCAT 490
DB 325 GTTGTGATTAACCTTCCCTCTATACATCAGTATCTCAAGTCGATACCAATATCCAT 384

QY 491 AAGAAGTGGTATATTTGTGAAAAAATAAAGTGGTATCTGGTATATACAATACCA 550
DB 385 AAGAAGTGGTATATTTGTGAAAAAATAAAGTGGTATCTGGTATATACAATACCA 444

QY 551 CGGTCTCGAATTCGCTCAACAATTTCTAGAGAAATGGAAGTGTCTCTTTGGTTTTATT 610
DB 445 CGGTCTCGAATTCGCTCAACAATTTCTAGAGAAATGGAAGTGTCTCTTTGGTTTTATT 504

QY 611 TTATTTCTTAATACATCTCTATATTTTAAACAACCTTCGATGTCTCGCTTAAATTTTCAAT 670
DB 505 TTATTTCTTAATACATCTCTATATTTTAAACAACCTTCGATGTCTCGCTTAAATTTTCAAT 564

QY 671 GTCCCTAAATTTCTTAATACATAAATCGTAAAGAAAAATTCGTGGAAGCCACAGGACATG 730
DB 565 GTCCCTAAATTTCTTAATACATAAATCGTAAAGAAAAATTCGTGGAAGCCACAGGACATG 624

QY 731 CATAGGCGAGTGTAGTTACCTTTTAAACCAATCAAAAATATATTAATAGAAAAGAACTTC 790
DB 625 CATAGGCGAGTGTAGTTACCTTTTAAACCAATCAAAAATATATTAATAGAAAAGAACTTC 684

QY 791 CTAAAAAGAAACAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGAGAGAAAAACG 850
DB 685 CTAAAAAGAAACAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGAGAGAAAAACG 744

QY 851 TATGGCCGAGTCTGTAAACAAGGACGTCCCGACCACTCGGAGACGGCGAGACGCTGAC 910
DB 745 TATGGCCGAGTCTGTAAACAAGGACGTCCCGACCACTCGGAGACGGCGAGACGCTGAC 804

QY 911 TGAATTTTCTTTTCTTTCTTAAAGACGTTGTTTGTGCTTACAGGGTCAAAACCA 970
DB 805 TGAATTTTCTTTTCTTTCTTAAAGACGTTGTTTGTGCTTACAGGGTCAAAACCA 864

QY 971 TATCCAATTTCTTGCCTATTATATATAAATAAAGATCCCTCTTGTGCTTTGTCTTTA 1030
DB 865 TATCCAATTTCTTGCCTATTATATATAAATAAAGATCCCTCTTGTGCTTTGTCTTTA 924

QY 1031 TTCTGTATATATAATCTAACTTAAATTTAGTTCTTAAATAATATATATGCTCCTACCTATGTTTC 1090
DB 925 TTCTGTATATATAATCTAACTTAAATTTAGTTCTTAAATAATATATATGCTCCTACCTATGTTTC 984

QY 1091 TACTGACCTCAGTCCCTAGTTAGCTATATGAGACATATGTGAAATGACGCCAAAAATTTG 1150
DB 985 TACTGACCTCAGTCCCTAGTTAGCTATATGAGACATATGTGAAATGACGCCAAAAATTTG 1044

QY 1151 AAGAGTTCCCTTCTTGCAACTAACTCTTACTCTTACTCATTTGAGCTATGTTAAATATGA 1210
DB 1045 AAGAGTTCCCTTCTTGCAACTAACTCTTACTCTTACTCATTTGAGCTATGTTAAATATGA 1104

QY 1211 ATGTTGGCACTCTCGTATTAAATATATGCACTGACCTAGATAAAAAAATCATGATAGACA 1270
DB 1105 ATGTTGGCACTCTCGTATTAAATATATGCACTGACCTAGATAAAAAAATCATGATAGACA 1164

QY 1271 TTTAGTTTAAAACTTGAAATGTTATTGAACTCTTTGGATTAAGTGGATTTGTTGATGGA 1330
DB 1165 TTTAGTTTAAAACTTGAAATGTTATTGAACTCTTTGGATTAAGTGGATTTGTTGATGGA 1224

QY 1331 TTAAATTTGAAGATATTTATATATTTGAAGATGTTTATATATATATTAGAGTTTATATAGCA 1390
DB 1225 TTAAATTTGAAGATATTTATATATTTGAAGATGTTTATATATATATTAGAGTTTATATAGCA 1284

QY 1391 GAAAAATATTGATGATGTTGTCCTTTTGTAGTTACTCTTTTGTGCGTAGTCTCTTT 1450
DB 1285 GAAAAATATTGATGATGTTGTCCTTTTGTAGTTACTCTTTTGTGCGTAGTCTCTTT 1344

QY 1451 CTCTCATCTCTCTATGAAGAAAAATCCAAAATAGTTTAAAGAAAAATTTTGTGTAATTCAT 1510
DB 1345 CTCTCATCTCTCTATGAAGAAAAATCCAAAATAGTTTAAAGAAAAATTTTGTGTAATTCAT 1404

QY 1511 AGTCTTTTTCGTAAACACACAGTTCTATGATAGCTATCGTCATCATATCTCTCTTTGCAACAA 1570
DB 1405 AGTCTTTTTCGTAAACACACAGTTCTATGATAGCTATCGTCATCATATCTCTCTTTGCAACAA 1464

QY 1571 CAAAAAGATCGTTTTTGTAAAAATTTAGTAGGCGACCTAAAGTCGTCAATTTGTTGTCCTGT 1630
DB 1465 CAAAAAGATCGTTTTTGTAAAAATTTAGTAGGCGACCTAAAGTCGTCAATTTGTTGTCCTGT 1524

QY 1631 CGAAATCTAGCGTTCTGTGATCCAAATTAAGTTGTTGATTCGAGCTCCAAAGATTATA 1690
DB 1525 CGAAATCTAGCGTTCTGTGATCCAAATTAAGTTGTTGATTCGAGCTCCAAAGATTATA 1584

QY 1691 ATCTTTTGTAGATGGTGCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1750
DB 1585 ATCTTTTGTAGATGGTGCATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATT 1644

QY 1751 TCTAACATATACGTCTTTTGTAGGCTCTGCGTCTTTTGTAGACCCACCCCTTTGCTA 1810
DB 1645 TCTAACATATACGTCTTTTGTAGGCTCTGCGTCTTTTGTAGACCCACCCCTTTGCTA 1704

QY 1811 ATGTTTGTGACCTTAGACAAATCCAAATAGCTTTACGTAGTGCAGAGTTGCAACAAA 1870
DB 1705 ATGTTTGTGACCTTAGACAAATCCAAATAGCTTTACGTAGTGCAGAGTTGCAACAAA 1764

QY 1871 TGGTCCAAATATAATTTTAAATTTGGCCACAAACCAACATTTTACAAACAAATTTCAACAAA 1930
DB 1765 TGGTCCAAATATAATTTTAAATTTGGCCACAAACCAACATTTTACAAACAAATTTCAACAAA 1824

QY 1931 CATGCACTGTTTCAAAATTTTATTTATTCATGCGGTTATTTGTTCAATTTGTAATTTCTG 1990
DB 1825 CATGCACTGTTTCAAAATTTTATTTATTCATGCGGTTATTTGTTCAATTTGTAATTTCTG 1884

QY 1991 TTTAACTCTACGCAAGATTTTAAATTTTCAAGAGAGAACATTTTGTATATAATAAACA 2050
DB 1885 TTTAACTCTACGCAAGATTTTAAATTTTCAAGAGAGAACATTTTGTATATAATAAACA 1944

QY 2051 TTTTATGGAAACCCCGTTTAAGCTCGATGATTTTGTAGTTTGTAGTTTGTGTTGTTGTA 2110

Db 1945 TTTTATGGAACACCGGTTAAGCTCGATGATTTTGGATTTTGTGCGTTTGTGAA 2004
Qy 2111 ATCATTAAAGACCTACATTTGATCCCTCATTTAATAATAGGAATCAAAATGATG 2170
Db 2005 ATCATTAAAGACCTACATTTGATCCCTCATTTAATAATAGGAATCAAAATGATG 2064
Qy 2171 ATTAAGTTTCAACAAAGAGCTCTCTTATGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2230
Db 2065 ATTAAGTTTCAACAAAGAGCTCTCTTATGCTATTAAGAGTCAGACGCAAGGATGACCGGG 2124
Qy 2231 GTCATTAAAGAGCTCTTATTAATCAACATTAATCTCCACTAATGCTTAATCAATCAG 2284
Db 2125 GTCATTAAAGAGCTCTTATTAATCAACATTAATCTCCACTAATGCTTAATCAATCAG 2178

RESULT 5
AX063409 LOCUS 2149 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 1 from Patent WO0100833.
ACCESSION AX063409
VERSION AX063409.1 GI:12541197
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Hoffmann, B., Mollier, P. and Pelletier, G.
Promoter expressed specifically in the cells of plant roots,
recombinant vectors and host cells comprising same and transgenic
plants obtained
Patent: WO 0100833-A 1 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
JOURNAL Location/Qualifiers
FEATURES
source
1..2149
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN
Query Match 48.7%; Score 2149; DB 6; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 136 GTGCAATTGTGATATATTGTAAGCAATCTGAAAGTAAGTGGATATATAAACAACG 195
Db 1 GTGCAATTGTGATATATTGTAAGCAATCTGAAAGTAAGTGGATATATAAACAACG 60
Qy 196 GCGAAGTACAGTTCTACCTTTTGTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 255
Db 61 GCGAAGTACAGTTCTACCTTTTGTGGCATGGAACCATGTTTGTAGGATTTACTTTGT 120
Qy 256 AATTCCTGAATCTTTCAATTCCTGAATTGATATTTACATTTTATCAAAAAAGTACA 315
Db 121 AATTCCTGAATCTTTCAATTCCTGAATTGATATTTACATTTTATCAAAAAAGTACA 180
Qy 316 AGTTCTACACAGACAGAGTTAAACAACCTTGCTGTCGAATGCTAATTTAAAGCCTAA 375
Db 181 AGTTCTACACAGACAGAGTTAAACAACCTTGCTGTCGAATGCTAATTTAAAGCCTAA 240
Qy 376 TCTTATGATTTCCCTTTTCTTCCAGATATATATGATTTGATATGACCCCATTTGTTG 435
Db 241 TCTTATGATTTCCCTTTTCTTCCAGATATATATGATTTGATATGACCCCATTTGTTG 300
Qy 436 TCATTAACCTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
Db 301 TCATTAACCTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
Qy 496 GTGGTATATTGTGAAAAAAGAAAAAAGTGGTATCTGGGTATATACATACACAGGTC 555
Db 361 GTGGTATATTGTGAAAAAAGAAAAAAGTGGTATCTGGGTATATACATACACAGGTC 420

Qy 556 TCGAATTCCTCAACAAATTTCTAGGAGAAAAATGACGTCTCTTTGGTTTATTTATTT 615
Db 421 TCGAATTCCTCAACAAATTTCTAGGAGAAAAATGACGTCTCTTTGGTTTATTTATTT 480
Qy 616 CTTAATAACATACCTCTATATTTTAAACACTTTCGATGTCCTCGCTTAAATTTTGAATGTGC 675
Db 481 CTTAATAACATACCTCTATATTTTAAACACTTTCGATGTCCTCGCTTAAATTTTGAATGTGC 540
Qy 676 TAAATTTCTCTAATCATTAATCGTAAAGAAATTCGTGAAAGCCACAGGACATCATAG 735
Db 541 TAAATTTCTCTAATCATTAATCGTAAAGAAATTCGTGAAAGCCACAGGACATCATAG 600
Qy 736 GGCAGTGTAGTTACCTTTTAAACCATCAAAAATATATTAATAGAAAGAAAACTTCTCTAAA 795
Db 601 GGCAGTGTAGTTACCTTTTAAACCATCAAAAATATATTAATAGAAAGAAAACTTCTCTAAA 660
Qy 796 AGAACAAATTTAATAAGTGGATATAAAAAAGATAGAAGGTAGGCAAGAAAAACGTATGG 855
Db 661 AGAACAAATTTAATAAGTGGATATAAAAAAGATAGAAGGTAGGCAAGAAAAACGTATGG 720
Qy 856 CCGGACCTCGTAAACAAGGAGCGTCCGACCACTCGGAGACGCGGAGAGCGCTGACTGATT 915
Db 721 CCGGACCTCGTAAACAAGGAGCGTCCGACCACTCGGAGACGCGGAGAGCGCTGACTGATT 780
Qy 916 TTTTCTTTTCTTTTCTTAAAGAAAGCTTTTCTGCTTTTACAAAGGTCAAAAACCATATCC 975
Db 781 TTTTCTTTTCTTTTCTTAAAGAAAGCTTTTCTGCTTTTACAAAGGTCAAAAACCATATCC 840
Qy 976 AATGTTTCTGCTATTTATATATAAATAAAGATCCCTCTTGTGCTTTGTCTTTATTCGT 1035
Db 841 AATGTTTCTGCTATTTATATATAAATAAAGATCCCTCTTGTGCTTTGTCTTTATTCGT 900
Qy 1036 GATATATATCTAATCTTAAATAGTCTTAAATAATATATATGCTACCTATGTTTCTACTG 1095
Db 901 GATATATATCTAATCTTAAATAGTCTTAAATAATATATATGCTACCTATGTTTCTACTG 960
Qy 1096 ACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAATACGCGCCCAAAAATTTGAAGAG 1155
Db 961 ACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAATACGCGCCCAAAAATTTGAAGAG 1020
Qy 1156 TTCTCTCTCTGCAACCTAACTCTTATCTTATCTCATGAGCTATGTTTAAATATTTGAATGTT 1215
Db 1021 TTCTCTCTCTGCAACCTAACTCTTATCTTATCTCATGAGCTATGTTTAAATATTTGAATGTT 1080
Qy 1216 GGCACCTCTGATTTAAATATGTCACCTAGATAAAAAAACAATGATAGACATTTAG 1275
Db 1081 GGCACCTCTGATTTAAATATGTCACCTAGATAAAAAAACAATGATAGACATTTAG 1140
Qy 1276 TTTAAACCTTGAAATGTTTATTTGAACTCTTTGGATTAGCTGGATTGTTGTATGGAATTTAAA 1335
Db 1141 TTTAAACCTTGAAATGTTTATTTGAACTCTTTGGATTAGCTGGATTGTTGTATGGAATTTAAA 1200
Qy 1336 TTTTGAAGATATTTATATATTTGAAGATGTTTATATATATATAGATTTTATATAGCAAAAA 1395
Db 1201 TTTTGAAGATATTTATATATTTGAAGATGTTTATATATATATAGATTTTATATAGCAAAAA 1260
Qy 1396 TATTGATCTGATGTTGTCCTTTTGTAGTACTCTTTTGTGTTGTTGTTGTTGTTCTCTCT 1455
Db 1261 TATTGATGATGATGTTGTCCTTTTGTAGTACTCTTTTGTGTTGTTGTTGTTGTTCTCTCT 1320
Qy 1456 CATCTCTCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTTCAATGCT 1515
Db 1321 CATCTCTCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTTCAATGCT 1380
Qy 1516 TTTTGTAAACACAGTTCTATGATGATCTGTCATCATATTCCTCTTTTGCACAAACAAAA 1575
Db 1381 TTTTGTAAACACAGTTCTATGATGATCTGTCATCATATTCCTCTTTTGCACAAACAAAA 1440
Qy 1576 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTTAAGTCTGATTTGTTGTTCTGTCGAAA 1635
Db 1441 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTTAAGTCTGATTTGTTGTTCTGTCGAAA 1500

Qy	1636	TCTAGCGTCTCTGTCATCCACAAATAAGTGTGTTGATTCCAGCTTCCAGATTATAAATCTT	1695	Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ds	1501	TCTAGCGTCTCTGTCATCCACAAATAAGTGTGTTGATTCCAGCTTCCAGATTATAAATCTT	1560	
Qy	1696	TTTATAGTGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA	1755	
Ds	1561	TTTATAGTGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA	1620	
Qy	1756	CATATAGCTCTGTTGTTTGGTAGGCTCTGGCTCTTTTGGAGCAACCCCTTCTGTAATGTT	1815	
Ds	1621	CATATAGCTCTGTTGTTTGGTAGGCTCTGGCTCTTTTGGAGCAACCCCTTCTGTAATGTT	1680	
Qy	1816	TTGTTGACCTTAGACCAATCCATAATAGTTTACGTGAGTGCAGTGCACCAATGCTC	1875	
Ds	1681	TTGTTGACCTTAGACCAATCCATAATAGTTTACGTGAGTGCAGTGCACCAATGCTC	1740	
Qy	1876	CAATATAAATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTTCAACAAACATGC	1935	
Ds	1741	CAATATAAATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTTCAACAAACATGC	1800	
Qy	1936	ATCGTTTCAAAATTTTAAATTTCAATGCGGTTATTTGTTCAATGTTAAATTTCTGTTAA	1995	
Ds	1801	ATCGTTTCAAAATTTTAAATTTCAATGCGGTTATTTGTTCAATGTTAAATTTCTGTTAA	1860	
Qy	1996	CTCACTGACGAATTTTAAATTTTCAAGAGAAACATTTTGTATATAAATAACATTTTA	2055	
Ds	1861	CTCACTGACGAATTTTAAATTTTCAAGAGAAACATTTTGTATATAAATAACATTTTA	1920	
Qy	2056	TGGAACCAACCGTTTAAAGTCGATGATTTTGAATTTTGAATTTTGTGTTTGTGAAATCAT	2115	
Ds	1921	TGGAACCAACCGTTTAAAGTCGATGATTTTGAATTTTGAATTTTGTGTTTGTGAAATCAT	1980	
Qy	2116	TAAAGCACTACATTTGATCCCTCATTTCTTAAATTAGGAATCAACATGATGATTA	2175	
Ds	1981	TAAAGCACTACATTTGATCCCTCATTTCTTAAATTAGGAATCAACATGATGATTA	2040	
Qy	2176	GTTTCAACAAAGAGCTCTTATGCTATTAAGCTTAAAGCTCAGACGCAAGGATGACCGGGTCA	2235	
Ds	2041	GTTTCAACAAAGAGCTCTTATGCTATTAAGCTTAAAGCTCAGACGCAAGGATGACCGGGTCA	2100	
Qy	2236	TAAAGCTCTTATATCAACCATTTACTTCACTAATTTGCTAATTAATCAG 2284		
Ds	2101	TAAAGCTCTTATATCAACCATTTACTTCACTAATTTGCTAATTAATCAG 2149		
RESULT 6				
AX063410				
LOCUS				
DEFINITION				
Sequence 2 from Patent WO0100833.				
ACCESSION				
AX063410				
VERSION				
AX063410.1 GI:12541198				
KEYWORDS				
synthetic construct				
SOURCE				
synthetic construct				
ORGANISM				
artificial sequences.				
REFERENCE				
1				
Hoffmann,B., Mollier,P. and Pelletier,G.				
AUTHORS				
TITLE				
Promoter expressed specifically in the cells of plant roots,				
recombinant vectors and host cells comprising same and transgenic				
plants obtained				
Patent: WO 0100833-A 2 04-JAN-2001;				
JOURNAL				
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)				
FEATURES				
source				
1..4280				
/organism="synthetic construct"				
/mol_type="unassigned DNA"				
/db_xref="taxon:32630"				
/note="Construction promoteur + sequence codante du g				
gus"				
ORIGIN				
Query Match 48.7%; Score 2149; DB 6; Length 4280;				
Best local Similarity 100.0%; Pred. No. 0;				


```
1216 GGCACTCTCGTATTAAATATGCCAGTTGCGACCTAGATAAAAAACATGATAGACATTTAG 1275
1081 GGCACCTCTCGTATTAAATATGCCAGTTGCGACCTAGATAAAAAACATGATAGACATTTAG 1140
1276 TTTAAACCTTGAAATGTTATTGAACTCTTTGGATAGCTGGATGTTGTTGATGATATAA 1335
1141 TTTAAACCTTGAAATGTTATTGAACTCTTTGGATAGCTGGATGTTGTTGATGATATAA 1200
1336 TTTTGAAGATATTATATATTGAAGATGTTTATATATATTAGATTATATATAGCAGAAAA 1395
1201 TTTTGAAGATATTATATATTGAAGATGTTTATATATATTAGATTATATATAGCAGAAAA 1260
1396 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTGCGTAGTCCCTTTCCT 1455
1261 TATTGATGTAGATGTTGTCCTTTTGTAGTTACTCTTTTTTGTGCGTAGTCCCTTTCCT 1320
1456 CATCTCCTATGAAGAAAAATCCAAATAGTTTAAAGAAAAATTTTGTGTAATTCATAGTCT 1515
1321 CATCTCCTATGAAGAAAAATCCAAATAGTTTAAAGAAAAATTTTGTGTAATTCATAGTCT 1380
1516 TTTTCGTAACACACAGTCTCATGTAGCTATCGTCATCATATTCCTCTTTTGGCAACAAAA 1575
1381 TTTTCGTAACACACAGTCTCATGTAGCTATCGTCATCATATTCCTCTTTTGGCAACAAAA 1440
1576 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAAGTCGTCAATTTGTTGTCCTGCGAAA 1635
1441 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAAGTCGTCAATTTGTTGTCCTGCGAAA 1500
1636 TCTAGCGTTCTGTCATCCACAAAAATAGTTGTTTGAATCGAGCTTCCAAGATTAATCTT 1695
1501 TCTAGCGTTCTGTCATCCACAAAAATAGTTGTTTGAATCGAGCTTCCAAGATTAATCTT 1560
1696 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1755
1561 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATAATTTCTAA 1620
1756 CATATACGTCCTGTTTGTGGTAGGCTCTGGCTCTTTTGGAGCCACCCCTTTCGTAATGTT 1815
1621 CATATACGTCCTGTTTGTGGTAGGCTCTGGCTCTTTTGGAGCCACCCCTTTCGTAATGTT 1680
1816 TTGTTGACCTTAGACAAATCCATATAGTTTACGTTAGTGAAGTGCACCAAAATGGTC 1875
1681 TTGTTGACCTTAGACAAATCCATATAGTTTACGTTAGTGAAGTGCACCAAAATGGTC 1740
1876 CAAATATAAATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTTCAACAAATGC 1935
1741 CAAATATAAATTTAAATTTGGCCACAAACAAACATTTTACAAACAAATTTCAACAAATGC 1800
1936 ATCGTTTCAAATTTTATTTCAATGGCGTTATTTGTTCAATGTAATATTTCTGTTTAA 1995
1801 ATCGTTTCAAATTTTATTTCAATGGCGTTATTTGTTCAATGTAATATTTCTGTTTAA 1860
1996 CTCACTGACGAATTTTAAATTTTCAAGAAAGACATTTTGTATATAAATTAACATTTTA 2055
1861 CTCACTGACGAATTTTAAATTTTCAAGAAAGACATTTTGTATATAAATTAACATTTTA 1920
2056 TGGAAACACCGGTTAAGTCCGATGATTTGAGTTTGTAGTTTGTGCGTTTGTGAAATCAT 2115
1921 TGGAAACACCGGTTAAGTCCGATGATTTGAGTTTGTAGTTTGTGCGTTTGTGAAATCAT 1980
2116 TAAACGACTACATTTGATCCCTCAATTTCTTAAATTAATPAGGAATCAACATGATGATTA 2175
1981 TAAACGACTACATTTGATCCCTCAATTTCTTAAATTAATPAGGAATCAACATGATGATTA 2040
2176 GTTACCAACAGACGTCTCTTATGCGTATTAAGAGTCAGACCAAGGATGACCGGGGTCAT 2235
2041 GTTACCAACAGACGTCTCTTATGCGTATTAAGAGTCAGACCAAGGATGACCGGGGTCAT 2100
2236 TAAGACGTCCTTATATCAACCAATTAATCTCCATTAATTCGTAATTAATCAG 2284
2101 TAAGACGTCCTTATATCAACCAATTAATCTCCATTAATTCGTAATTAATCAG 2149
```

```
RESULT 7
AF242313 LOCUS 1741 bp DNA linear PLN 06-AUG-2000
DEFINITION Arabidopsis thaliana unknown sequence.
ACCESSION AF242313
VERSION AF242313.1 GI:9719364
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1 (bases 1 to 1741)
AUTHORS Mollier, P., Hoffmann, B., Pelletier, G. and Chwetzoff, S.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2000) Genetique, INRA, Route de St Cyr,
Versailles 78026, France
FEATURES
source
1. .1741
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="I"
/map="NIAI-agp64"
/ecotype="Massilewskija"
<1. .>1741
/misc_feature
/note="contains root-specific promoter"
ORIGIN
Query Match 33.7%; Score 1486; DB 8; Length 1741;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1736; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 544 AATACGAGGTCTCGAATTCGCTCAACAATTTCTAGGAGAAAAATGGAGCTGCTCTTTGG 603
Db 1 AATACGAGGTCTCGAATTCGCTCAACAATTTCTAGGAGAAAAATGGAGCTGCTCTTTGG 60
Qy 604 TTTTATTTTATTTCTTAATAACATACACTCTATATTTTAAACACTTCGATGCTCGCTTAAT 663
Db 61 TTTTATTTTATTTCTTAATAACATACACTCTATATTTTAAACACTTCGATGCTCGCTTAAT 120
Qy 664 TTCGAATGTGCTAAATTTCTCTAATCAATAAATCGTAAAGAAAAATTCGCGAAGCCACAG 723
Db 121 TTCGAATGTGCTAAATTTCTCTAATCAATAAATCGTAAAGAAAAATTCGCGAAGCCACAG 180
Qy 724 GGACATGCATAGGCGAGTAGTTACCTTTAAACCATCAAAAATATATTAATAGAAAGG 783
Db 181 GGACATGCATAGGCGAGTAGTTACCTTTAAACCATCAAAAATATATTAATAGAAAGG 240
Qy 784 AAACCTCTTAAAGAAACAAATTTAATAAGTGATATAAAGATAAAGAGTAGGCAGAA 843
Db 241 AAACCTCTTAAAGAAACAAATTTAATAAGTGATATAAAGATAAAGAGTAGGCAGAA 300
Qy 844 GAAACGATGTGGCGCGACCTCGTAAACAGGAGCGTCCGACCACTGCGGAGACGGCGAGA 903
Db 301 GAAACGATGTGGCGCGACCTCGTAAACAGGAGCGTCCGACCACTGCGGAGACGGCGAGA 360
Qy 904 CGCTGACGTGATTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 963
Db 361 CGCTGACGTGATTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 420
Qy 964 AAACCATATATCAATTTCTGCTCTATTTATATATACTAAAGATCCCTCTTTCTGCTTT 1023
Db 421 AAACCATATATCAATTTCTGCTCTATTTATATATACTAAAGATCCCTCTTTCTGCTTT 480
Qy 1024 GTCTTTATTCGATATATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1083
Db 481 GTCTTTATTCGATATATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
Qy 1084 ATGTTTCTACTGACCTCAGTCCCTTAGTAGCTATATGGACATATGTGAAATGACGCCCA 1143
Db 541 ATGTTTCTACTGACCTCAGTCCCTTAGTAGCTATATGGACATATGTGAAATGACGCCCA 600
```

QY	2224	GACCGGGTCATTAAAGCGCTCTTATATTCAACCATTTACTCCACTAATTCGTAATTTAATCA	2283
Db	1681	GACCGGGTCATTAAAGCGCTCTTATATTCAACCATTTACTCCACTAATTCGTAATTTAATCA	1740
QY	2284	G 2284	
Db	1741	G 1741	
RESULT 8			
ATH553603/c			
LOCUS	ATH553603	317 bp	DNA linear PLN 29-MAR-2003
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence, left border, clone 362F10.		
ACCESSION	AJ553603		
VERSION	AJ553603.1	GI:29369764	
KEYWORDS	left border; T-DNA flanking sequence.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1		
AUTHORS	Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny, A.		
TITLE	T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites		
JOURNAL	EMBO Rep. 3 (12), 1152-1157 (2002)		
MEDLINE	22363535		
PUBMED	12446565		
REFERENCE	2 (bases 1 to 317)		
AUTHORS	Balzerque, S.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE		
COMMENT	PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/ . This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).		
FEATURES	Location/Qualifiers		
source	1..317		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/cultivar="Wassillewskija"		
	/db_xref="taxon:3702"		
	/clone="362F10"		
	/clone_lib="Arabidopsis thaliana T-DNA insertion lines"		
misc_feature	1..317		
	/note="T-DNA flanking sequence left border"		
ORIGIN			
Query Match	4.6%; Score 202; DB 8; Length 317;		
Best Local Similarity	99.6%; Pred. No. 6.2e-91;		
Matches	252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1888	AAATTTGGCCACAAAACAACATTTTACAAACAATTTCAACAACATGTCATCGTTTCAAAAT	1947
Db	317	AAATTTGGCCACAAAACAACATTTTACAAACAATTTCAACAACATGTCATCGTTTCAAAAT	258
QY	1948	TTTATTATTCAATGGCGTTATTGTTTCATTGTTAAATATTCTGTTAACTCACTGACGAA	2007
Db	257	TTTATTATTCAATGGCGTTATTGTTTCATTGTTAAATATTCTGTTAACTCACTGACGAA	198
QY	2008	TTTTTTAAATTTTCAAGAAGAACATTTTGTATATAATAAATTTTATGGAACACCACCGG	2067


```

/protein_id="AAQ22655.1"
/db_xref="GI:33589778"
/translation="MLGLRSATTLFDISQSLRNVTFHGLRVQGIHVGNAEVPPNNKP
LKTGLQEVGIGRKSQVHIGLITNKILADLTGKELIDREEVGHQHGDLRRRV
GSEIQRLVEVDCVGRSRHRHGLPCRQRTSTNARTKKGAVALAGKKAPRK"

ORIGIN
Query Match      1.6%; Score 69; DB 8; Length 465;
Best Local Similarity 100.0%; Pred. No. 4.6e-23;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3738 ATGTTGGTTTACGAAGATCTGCGACGACCTTGTTGCACATCAGCCAGTCTCTGCTTCGT 3797
Db 1 ATGTTGGTTTACGAAGATCTGCGACGACCTTGTTGCACATCAGCCAGTCTCTGCTTCGT 60
QY 3798 AATGTTACG 3806
Db 61 AATGTTACG 69

RESULT 11
AL773135
LOCUS Arabidopsis thaliana transposon insertion STS SM_3.1249, sequence
DEFINITION tagged site.
ACCESSION AL773135
VERSION AL773135.1 GI:21535322
KEYWORDS STS; STS, sequence tagged site.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B.,
Murphy,G., Lingham,S., LeGrys,C., Jones,J.D.G. and Bevan,M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 295)
AUTHORS Clarke,J.H.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-2002) Clarke J.H., John Innes Centre, Colney
Lane, Norwich, NR4 7UJ, UK
COMMENT AT denotes an activation tag dissociation transposon within a
single line, ET an enhancer trap dissociation transposon, GT a gene
trap dissociation transposon, MT a mis-expression enhancer trap
dissociation transposon, SM a defective suppressor mutator
transposon. 3 denotes a sequence derived from the 3' end of the
transposon, _5 denotes a sequence derived from the 5' end of the
transposon. BESRC GARNET, ATIS project
On-line seed stock requests: http://nasc.nott.ac.uk/ NASC stock
code: N55909.
FEATURES
source
Location/Qualifiers
1..295
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia-0 NASC stock code N1092"
/db_xref="taxon:3702"
/clone="AC012193"
/note="Derived from superpool 12.77 NASC code Unknown"
STS
1..295
/standard_name="SM_3.1249"

ORIGIN
Query Match      1.5%; Score 68; DB 11; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.5e-22;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 ACATCAGTATCTCAAAGTCGAATACCAATATCCATAAGAGTGCTATATTGTGAAAAAA 515
Db 1 ACATCAGTATCTCAAAGTCGAATACCAATATCCATAAGAGTGCTATATTGTGAAAAAA 60
QY 516 AAAAAAAA 523
|||||||

```

```

Db 61 AAAAAAA 68

RESULT 12
AC007289/c
LOCUS Arabidopsis thaliana chromosome 2 clone F16J10 map RNS1, complete
DEFINITION sequence.
ACCESSION AC007289
VERSION AC007289.9 GI:20198243
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 43982)
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M.,
Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,
Nierman,W.C., Fraser,C.M. and Venter,J.C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 43982)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 43982)
AUTHORS Town,C.D. and Kaul,S.
JOURNAL Direct Submission
TITLE Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598749.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
source
Location/Qualifiers
1..43982
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="RNS1"
/clone="F16J10"
/misc_feature
1..5724
/note="overlap with BAC clone T20G20
(AC006220:104629..110352)."
```

```

mRNA
complement(join(<5500..5949,6031..>.6147))
/gene="At2g05430"
complement(join(<5500..5949,6031..>.6147))
/gene="At2g05430"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA015480.1"
/db_xref="GI:20198245"
/translation="MSKISEKSTLSHHGTAYDELDNDENVTKGWIFRSKGLAK
RVPEKHPETTWKFLCKSEFAKETVLTALLDIDMINLPQOSLFEAEKKAENTILD
EAAGFKLDWLKRLKEERVYKKAKNRTTRMRELDRLKQKLELSVLQEBMKKEQPE
AMDKPEYDLSLCLNEVTCDSSEPIK"
complement(7260..8700)
/gene="At2g05420"
/note="synonym: F16J10.3; predicted by genefinder"
complement(join(<7260..7640,7857..8077,8329..8510,
8591..>.8700))
/gene="At2g05420"
complement(join(7260..7640,7857..8077,8329..8510,
8591..8700))
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32895.1"
/db_xref="GI:4914358"
/translation="MGKQVDSKTTITWVIEFSSSQSAIHSDFQVGVGDKRWLKAYPK
FNEKSTLAYRANLALYLVANGSKSPFPIGWRHTKFLSLTVNOKSEKLSKLTSEQHW
GDHXTSRGPAPMPLTNLTNEGFLVNGELTLVAKVELEVGVGLDVSKKSPWMT
IDVNGFOVLPOIESVNRLLFAKHLDIASKRFPKPKMYMTAYNNVLLSLTKLTCQSPQDL
IYDKISGAALTYLREAGFKLDWLEKKHGEIKKKKBEASLKLQEMEXQIFNEAQ
9262..9302
/rpt_family="AT rich"
complement(9910..9970)
/rpt_family="AT_rich"
complement(10000..10019)
/rpt_family="TAAA)n"
10066..10101
/rpt_family="AT rich"
complement(14048..15111)
/gene="At2g05410"
/note="synonym: F16J10.4; predicted by genscan"
complement(join(<14048..14464,14624..14847,14955..>.15111))
/gene="At2g05410"
complement(join(14048..14464,14624..14847,14955..>.15111))
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32896.1"
/db_xref="GI:4914359"
/translation="MAYPKGINKAHDSEFLNVDPNBSLPTGWRHRAKVFSVLNQG
SEKLSQRVTOHWFOKATWGFPMVITHTEINAKMGFLVNGELKVVAKIEVLEVGK
LDVSKESPIMKTDVNGFOVLPSQVDSVKFLFKNLDIVSKFLKPNLYLTACWNLL
LSLETLCQSFOELNSDGLSAGVALAYITETGLKDWLEKKLDELKELKKKRSCLV
RLREMDQIQPFKRCLEDIQISKEKBELLAAAREPLSYDDIDNNV"
complement(15873..15972)
/rpt_family="AT rich"
complement(16262..17096)
/gene="At2g05400"
/note="synonym: F16J10.5; predicted by genscan"
complement(join(<16262..16672,16854..>.17096))
/gene="At2g05400"
complement(join(16262..16672,16854..>.17096))
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD32897.1"
/db_xref="GI:4914360"
/translation="MFCDSAETQWFKDKNAPDWGFTEMLSLTELHAERGLRNGELTV
VAKVEVLEWVGKLDVSESSPVIEIDVNGFOVLPSQVEYAKSLFERHLDIASKRPK
NPLYKTAYNNVLSLTQICQSFQELSDNDLSDAGALAYLREAGFELDWLKELNEV
KEKKKBEACAEIQQMDHVKPLKKYLLDLLEAQIDKKKAEALLAARPLSLNDNNV"
complement(17875..21892)

```

```

mRNA
complement(join(<17875..19791,19831..>.20615,20671..>.21892))
/gene="At2g05390"
complement(join(17875..19791,19831..>.20615,20671..>.21892))
/gene="At2g05390"
/codon_start=1
/product="putative retroelement pol polyprotein"
/protein_id="AAD32898.1"
/db_xref="GI:4914361"
/translation="MEATLRVHKVWETIDPGSDDMDEKNDMARALLFQSVPESTILQVG
XHKISKAWRAIKTRNLGAERVKEAKQLTMAEFDRLNMKDNIEIDFVGRISISITX
SESLGEEIESKIVKFKLSPRKKYIHTIAALEQILDLTITGFEDIVGRMKTVEDRV
CEDDSEPEEQKLYANSESSYTRGGRGRSGRGGGYQQRDQSKVICYRC
CDKHTYASECLDRLLKLIKAQEQOQNNEDDEIEISLMHVEVLYNERSVKEPFEAC
DNSVLDNGASNMTGNLWFSLKNEMITGKRVFGDSDRIDIKGSGTILVITKGGIRK
TLTDVYFIPDLKSNIIISLGQATEAGCDVRMKDDOLTLDHREGCLLRLATRSRNLKYV
DLNVNFKLQLEAATWVRKELVIGISNIPEKETCSCLLCKQARQFPFKATTYRAS
OVLEVHGDLCGPITQSTTAKRKVILYLLIDHTYRMMWMLLKESEAFKEDFKTKV
EQESGVKIKTFRDQGEFVSEQDFCAKEGINRHLTAPYTPQONGVVERNRTLLG
MTRSLKHKMPNYLWGEAVRHSYIIINRVGTRSLQNOTPYEVFKQKPNVHLRVFG
CIGYAKTEGPHRLKLDKSKMLVYLGTPEGSKAYRLDPTNRKIIKNWSSETRDIS
GTFSILTEGEGNGIQESDDILETEKNGEESSENESENEQOQIDAETOPSHA
TPPTLRSTRQVGNLYLDYVLMALIEQEVLLAINDEFPDPEKANKLEWDACK
EELISBKNKTWSLIDLVRKKYIGLKWVFKIKRNSDGSINKYIKARLVAKYGVORHG1
DYDFVFAHVARITIRVITIALAASNGWEVHLVDKTAFLHGELEDVVYVTPQEGTK
DNEGVKLKHALYGLQAPRAMTKNLKILQELNVFKSKEPSVYRQEEKLLIIVA
IYVDLLVTSGLDILCFKDMAGKTFMSDLGOLTYVYLGIEVLRKNGIILRQERYA
MKIIEAGMNCNPVLIPMAAGLECKAQEKICITERDYREMIICGLRYIVHTRPDL5Y
CVGWSYLOQPRSHGNALQVRYLYKGTWSHGLYIKRGFSGLVGVSDSHSDLLD
DGKSTAGHIFYLHQCPITWCQKQOVALSCEAFEMAATAAKQAILWDLFAEVCG
TTSEKVMIRVDNKSAIALTKNLVPHGRSKTHIRYHFIKRECVNNLVEVDHVPVGEQR
ADILTKPLGRIFKREMBRELGVQCVEDEDFKLKGEFVGISMKSIN"
complement(22479..22502)
/rpt_family="AT_rich"
complement(24570..25781)
/gene="At2g05380"
/note="synonym: F16J10.7; supported by cDNA:
gi.15146251.gb AY049267.1"
complement(join(<24570..25058,25678..>.25781))
/gene="At2g05380"
complement(join(24781..25058,25678..>.25750))
/gene="At2g05380"
/codon_start=1
/product="expressed protein"
/protein_id="AAD32899.2"
/db_xref="GI:20198244"
/translation="MASTLILLGLFAFLFIVSEMAAGTVKSESEETVKPEOHGGGF
GDNGGRYQGGGGHGGGGYQGGGRYQGGGRQGGGGYCYRHGCCYKGYHGGCSRC
CSYAGEAVQTQSGH"
complement(24878..24990)
/rpt_family="GGA)n"
complement(25118..25153)
/rpt_family="AT_rich"
complement(26299..26329)
/rpt_family="A)n"
complement(28917..28939)
/rpt_family="AT_rich"
28929..29834
repeat_region
Query Match 1.2%; Score 51; DB 8; Length 43982;
Best Local Similarity 100.0%; Pred. No. 5.9e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 215 CTTTTTTGGCATGGACCATGTTTITAGGATTTACTTTTGAATTCCTGAA 265
Db 34623 CTTTTTTGGCATGGACCATGTTTITAGGATTTACTTTTGAATTCCTGAA 34573

```

```

RESULT 13
AX347002/c AX347002 15046 bp DNA linear PAT 01-FEB-2002
LOCUS
DEFINITION Sequence 2073 from Patent WO0200928.

```

ACCESSION AX347002
VERSION AX347002.1 GI:18494888
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2073 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source Location/Qualifiers
1. .15046
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 0.6%; Score 26; DB 6; Length 15046;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3349 AAAAAATATATATAATAATAATAATAATAA 3374
|||||
Db 9895 AAAAAATATATATAATAATAATAATAATAA 9870

RESULT 14
AC144806
LOCUS Medicago truncatula clone mth2-34h22, complete sequence.
DEFINITION
AC144806
VERSION
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM

REFERENCE 1 (bases 1 to 124550)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Medicago truncatula BAC Clone mth2-34h22
JOURNAL
REFERENCE 2 (bases 1 to 124550)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 3 (bases 1 to 124550)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 4 (bases 1 to 124550)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 5 (bases 1 to 124550)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA
On Sep 10, 2003 this sequence version replaced gi:34536751.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
source Location/Qualifiers
1. .124550
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-34h22"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match 0.6%; Score 26; DB 8; Length 124550;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 ATATTGTGAAAAAAGT 526
|||||
Db 69422 ATATTGTGAAAAAAGT 69447

RESULT 15
AC142222/c
LOCUS Medicago truncatula clone mth2-27d20, WORKING DRAFT SEQUENCE, 4
DEFINITION
AC142222
VERSION
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM

REFERENCE 1 (bases 1 to 125992)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Medicago truncatula BAC Clone mth2-27d20
JOURNAL
REFERENCE 2 (bases 1 to 125992)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 3 (bases 1 to 125992)
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT
On May 31, 2003 this sequence version replaced gi:30698588.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4334: contig of 4334 bp in length

```
* 4335 4434: gap of unknown length
* 4435 29695: contig of 25261 bp in length
* 29696 29795: gap of unknown length
* 29796 69537: contig of 39742 bp in length
* 69538 69637: gap of unknown length
* 69638 125992: contig of 56355 bp in length.
```

FEATURES

```
source
1..125992
  Location/Qualifiers
    organism="Medicago truncatula"
    mol_type="genomic DNA"
    db_xref="taxon:3880"
    clone="mth2-27d20"
    clone_lib="Medicago truncatula BAC library H2"
```

ORIGIN

```
Query Match      0.6%; Score 26; DB 2; Length 125992;
Best Local Similarity 100.0%; Pred.No. 0.33;
Matches 26, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      501 ATATTGTGAAAAAAGT 526
        |||||
Db      62571 ATATTGTGAAAAAAGT 62546
```

Search completed: July 31, 2004, 15:50:55
Job time : 16512 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 08:20:44 ; Search time 1565 Seconds
(without alignments)
11979.104 Million cell updates/sec

Title: US-10-027-880-3

Perfect score: 4413

Sequence: 1 aagatcacagtgataaat.....tttcattgggtacagtcaca 4413

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4413	100.0	4413	4	AAF25318 Nucleotid
2	2154	48.8	4309	4	AAF25319 Nucleotid
3	2149	48.7	2149	4	AAF25316 Nucleotid
4	2149	48.7	4280	4	AAF25317 Nucleotid
5	80	1.8	598	3	AAC35408 Arabidops
c 6	26	0.6	15046	6	ABL34100 Human imm
7	25	0.6	243	5	ABV20030 Human pro
8	25	0.6	364	5	ABV49789 Human pro
9	25	0.6	2033	2	AAV40523 Homo sapi
10	24	0.5	388	8	ACH21196 Human adu
c 11	24	0.5	7752	6	ABL70224 Chemical
12	24	0.5	8964	6	ABL3535 Human imm
13	24	0.5	8964	6	ABK31373 Signal tr
14	24	0.5	8964	6	ABL70582 Chemical
15	24	0.5	8964	6	AS61277 Human gen
c 16	23	0.5	326	4	AAS38028 Novel hum
17	23	0.5	367	2	AAH84366 Drosophil
c 18	23	0.5	390	9	ADB55118 Toxicity
c 19	23	0.5	390	9	ADB49649 Primary r
20	23	0.5	399	4	AAI188708 Human pol
c 21	23	0.5	458	4	ABL35547 Human mus
c 22	23	0.5	458	7	ABX58535 cdna enco
23	23	0.5	976	4	AAL37396 Human mus

24	23	0.5	976	7	ABX60384	Abx60384	CDNA	enco
25	23	0.5	2292	5	AAS88155	Aas88155	DNA	enco
26	23	0.5	2501	6	ABZ82595	Abz82595	Human	sec
27	23	0.5	2532	9	ADB62172	Adb62172	Human	CDN
28	23	0.5	4636	2	AAZ30710	Aaz30710	Rat	neuro
29	23	0.5	4654	4	AAS46781	Aas46781	Tumour	su
30	23	0.5	4654	6	ABL34223	Abi34223	Human	imm
31	23	0.5	6311	6	ABL33963	Abi33963	Human	imm
32	23	0.5	110000	3	AAF22303	Continuation (3 of		
33	23	0.5	203654	7	ABX16034	Abx16034	Human	gen
34	22	0.5	181	4	AAI88838	Aai88838	Human	pol
35	22	0.5	182	6	ABL77052	Abi77052	Human	ova
c 36	22	0.5	241	5	ABV07853	Abv07853	Human	pro
37	22	0.5	256	6	ABL80025	Abi80025	Human	ova
38	22	0.5	294	8	ACH16395	Ach16395	Human	adu
c 39	22	0.5	373	5	ABV37777	Abv37777	Human	pro
40	22	0.5	375	4	AAL22740	Aal22740	Human	bre
41	22	0.5	391	6	ABV99309	Abv99309	Marine	sn
c 42	22	0.5	396	4	AAF94948	Aaf94948	Human	ova
c 43	22	0.5	396	6	ABL48898	Abi48898	Ovarian	c
c 44	22	0.5	396	6	ABT03215	Abt03215	Human	ova
45	22	0.5	397	4	AAL13874	Aal13874	Human	bre

ALIGNMENTS

RESULT 1

AAF25318

ID AAF25318 standard; DNA; 4413 BP.

XX

AC AAF25318;

XX

DT 30-APR-2001 (first entry)

XX

DE Nucleotide sequence of an Arabidopsis thaliana genomic fragment.

XX

KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.

XX

OS Arabidopsis thaliana.

XX

PN WO200100833-A1.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-FR001768.

XX

PR 25-JUN-1999; 99FR-00008185.

XX

(INRG) INRA INST NAT RECH AGRONOMIQUE.

XX

PI Hoffmann B, Mollier P, Pelletier G;

XX

DR WPI; 2001-102893/11.

XX

PT New constitutive plant promoter active specifically in roots, useful for
controlling expression of pest or drought resistance genes, and related
transgenic plants.

XX

XX Claim 3; Page 71-73; 92pp; French.

XX

CC The present sequence represents a genomic fragment of Arabidopsis
thaliana. The sequence contains a plant promoter that directs expression
of a selected sequence in root cells at all stages of development of a
plant. The plant promoter is used to control expression of genes in a
root-specific manner, especially genes that provide resistance to
parasites, pests (nematodes or fungi), water and salt stress, or alter
sugar content or nitrogen transport. Fragments of the promoter are useful
as probes or primers to detect or amplify at least part of the promoter

XX

SQ Sequence 4413 BP; 1427 A; 731 C; 702 G; 1553 T; 0 U; 0 Other;

Query Match 100.0%; Score 4413; DB 4; Length 4413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGATCCACAGTGAATAAATAAGAACCGATTCCGGTGATATGCAACTATATATAGAA	60
DB			
QY	1	AGATCCACAGTGAATAAATAAGAACCGATTCCGGTGATATGCAACTATATATAGAA	60
DB			
QY	61	ATTGAATACCTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATCTCTA	120
DB			
QY	61	ATTGAATACCTCTGATTCATCGCTTTGTATCAAGATCGAATCTCTAAAAACATATCTCTA	120
DB			
QY	121	TAATAAATATCTGAGTCGAATGTGTGATATATTGTAAAGCAATCTGAAAGAAATAGTGG	180
DB			
QY	121	TAATAAATATCTGAGTCGAATGTGTGATATATTGTAAAGCAATCTGAAAGAAATAGTGG	180
DB			
QY	181	ATATATAACAACCGCGGAAGTACAAAGTCTACCTTTTTCGCATGGAACCATGTTTT	240
DB			
QY	181	ATATATAACAACCGCGGAAGTACAAAGTCTACCTTTTTCGCATGGAACCATGTTTT	240
DB			
QY	241	TAGGATTTACTTTGTAACTCCTGAATCTTTCACTTCTGAAATGATATTTACATTTTAT	300
DB			
QY	241	TAGGATTTACTTTGTAACTCCTGAATCTTTCACTTCTGAAATGATATTTACATTTTAT	300
DB			
QY	301	CAAAAAAAGTACAGTTCTACCAAGACAGAGTTAAACAACTTGTTGTCAATGC	360
DB			
QY	301	CAAAAAAAGTACAGTTCTACCAAGACAGAGTTAAACAACTTGTTGTCAATGC	360
DB			
QY	361	TAATTTAAAGCCTAACTTATGATTTCCCTTTTCTCACGATATATCTGATATGATAT	420
DB			
QY	361	TAATTTAAAGCCTAACTTATGATTTCCCTTTTCTCACGATATATCTGATATGATAT	420
DB			
QY	421	GCACCCATTTGTTGTCAATTAACCTCCACTCTATACATCAGTATCTCAAAGTCGAATAA	480
DB			
QY	421	GCACCCATTTGTTGTCAATTAACCTCCACTCTATACATCAGTATCTCAAAGTCGAATAA	480
DB			
QY	481	CAATATCCAATAAGAGTGTATATTGTGAAAAAATAAATAAAGTGTATATCTGATATA	540
DB			
QY	481	CAATATCCAATAAGAGTGTATATTGTGAAAAAATAAATAAAGTGTATATCTGATATA	540
DB			
QY	541	TACAATACCAAGCTCTCGAATTCCTCAACAAATTTCTAGGAGAAATGGAAGTCTCTTT	600
DB			
QY	541	TACAATACCAAGCTCTCGAATTCCTCAACAAATTTCTAGGAGAAATGGAAGTCTCTTT	600
DB			
QY	601	TGGTTTATTTATTTCTTTAATAACATCTCTATATTTTAAACACTTCGATGTCTCGCTTA	660
DB			
QY	601	TGGTTTATTTATTTCTTTAATAACATCTCTATATTTTAAACACTTCGATGTCTCGCTTA	660
DB			
QY	661	AATTTCCAATGTGCTTAAATTTCTCTAATCAATAAGTAAAGAAATTCGTGGAAGCCA	720
DB			
QY	661	AATTTCCAATGTGCTTAAATTTCTCTAATCAATAAGTAAAGAAATTCGTGGAAGCCA	720
DB			
QY	721	CAGGACATGATAGGACAGTATGTTACCTTTTAAACCATCAAATAATATTAATAGAAA	780
DB			
QY	721	CAGGACATGATAGGACAGTATGTTACCTTTTAAACCATCAAATAATATTAATAGAAA	780
DB			
QY	781	AGGAAACTTCTTAAAGAACATTTAATAAGTGGATATAAATAAGATATAAGAGTAGGCA	840
DB			
QY	781	AGGAAACTTCTTAAAGAACATTTAATAAGTGGATATAAATAAGATATAAGAGTAGGCA	840
DB			
QY	841	GAAGAAACGATATGCGCGGACATCTGTAACAAGGAGCTCCGACCATCTCGGAGACGGCG	900
DB			
QY	841	GAAGAAACGATATGCGCGGACATCTGTAACAAGGAGCTCCGACCATCTCGGAGACGGCG	900
DB			
QY	901	AGACGCTGACTGATTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCGTCTTACAAGG	960
DB			
QY	901	AGACGCTGACTGATTTTCTTTTCTTTTCTTAAAGAACGTTGTTTCGTCTTACAAGG	960
DB			
QY	961	GTCAAAACCATATCCAATTTCTGCTTATTTATATACTAAAGTCCCTCTTGTGC	1020
DB			
QY	961	GTCAAAACCATATCCAATTTCTGCTTATTTATATACTAAAGTCCCTCTTGTGC	1020
DB			

QY	1021	TTTGTCTTTATTTCTGATATATAAATCTAACTTAANTTAGTTCTAAATAATATATATGCTTA	1080
DB			
QY	1021	TTTGTCTTTATTTCTGATATATAAATCTAACTTAANTTAGTTCTAAATAATATATATGCTTA	1080
DB			
QY	1081	CCATATGTTTCTACTGACCTCAGTCCCTAGTATATGACATATATGCAAAATGACGC	1140
DB			
QY	1081	CCATATGTTTCTACTGACCTCAGTCCCTAGTATATGACATATATGCAAAATGACGC	1140
DB			
QY	1141	CCAAAATTTGAAGAGTTCCTCTTCTCTGCAACTTATCTTACTCATTTGAGCTATGT	1200
DB			
QY	1141	CCAAAATTTGAAGAGTTCCTCTTCTCTGCAACTTATCTTACTCATTTGAGCTATGT	1200
DB			
QY	1201	TAAATATTGNACTGTTGGCACTCTCGTATTAAATATGCCAGTTCGACCTAGATAAAAAC	1260
DB			
QY	1201	TAAATATTGNACTGTTGGCACTCTCGTATTAAATATGCCAGTTCGACCTAGATAAAAAC	1260
DB			
QY	1261	ATCATAGACATTTAGTTTAAAACTTGAAATGTTATTGAACTCTTTGGATTAGCTGATT	1320
DB			
QY	1261	ATCATAGACATTTAGTTTAAAACTTGAAATGTTATTGAACTCTTTGGATTAGCTGATT	1320
DB			
QY	1321	GTTGTATGGAATTTAAATTTTGAAGATTTATATATTTGAAGATGTTTATATATATAGAT	1380
DB			
QY	1321	GTTGTATGGAATTTAAATTTTGAAGATTTATATATTTGAAGATGTTTATATATATAGAT	1380
DB			
QY	1381	TTATATAGCAGAAAAATATGATGATGATGTTGCTTTCTGTTAGTTACTCTTTTTTGTGC	1440
DB			
QY	1381	TTATATAGCAGAAAAATATGATGATGATGTTGCTTTCTGTTAGTTACTCTTTTTTGTGC	1440
DB			
QY	1441	GTAGTCTCTTCTCCTCATCTCTATGAAGAAAAATCCAAATAGTTTAAAGGAAATTTTG	1500
DB			
QY	1441	GTAGTCTCTTCTCCTCATCTCTATGAAGAAAAATCCAAATAGTTTAAAGGAAATTTTG	1500
DB			
QY	1501	TGTAATTCATAGTCTTTTTCGTAACACAGTTCCTATGTTAGTATCTGTCATCATATTCCTC	1560
DB			
QY	1501	TGTAATTCATAGTCTTTTTCGTAACACAGTTCCTATGTTAGTATCTGTCATCATATTCCTC	1560
DB			
QY	1561	TTTGGCAACACAAAAAGATCGTTTGTAAAAATTTAGTAGGCACCTAAAAGTCTCATTT	1620
DB			
QY	1561	TTTGGCAACACAAAAAGATCGTTTGTAAAAATTTAGTAGGCACCTAAAAGTCTCATTT	1620
DB			
QY	1621	GTTGTCTCTGCGAAATCTAGCTTCTGTCATCCACAAATTAAGTTGTTTGAATCGAGCTTC	1680
DB			
QY	1621	GTTGTCTCTGCGAAATCTAGCTTCTGTCATCCACAAATTAAGTTGTTTGAATCGAGCTTC	1680
DB			
QY	1681	CAGATATATATCTTTTGTAGTGGGTGATGAGATTTCTAACTTGTATACGAGTCTAT	1740
DB			
QY	1681	CAGATATATATCTTTTGTAGTGGGTGATGAGATTTCTAACTTGTATACGAGTCTAT	1740
DB			
QY	1741	CCATATAATTTCTAACATATACGTCCTGTTTGTGGTAGGCTCTGCGTCTTTTGAAGACCAC	1800
DB			
QY	1741	CCATATAATTTCTAACATATACGTCCTGTTTGTGGTAGGCTCTGCGTCTTTTGAAGACCAC	1800
DB			
QY	1801	CCCCTTGTCTAAGTTTGTGCACTTAGCAATCCATATAATACGTTACGTTGAGTCGAAGT	1860
DB			
QY	1801	CCCCTTGTCTAAGTTTGTGCACTTAGCAATCCATATAATACGTTACGTTGAGTCGAAGT	1860
DB			
QY	1861	TGCACCAAAATGTCCTCAATATATTAATTTGGCCACAAACCAACATTTTACAAACAA	1920
DB			
QY	1861	TGCACCAAAATGTCCTCAATATATTAATTTGGCCACAAACCAACATTTTACAAACAA	1920
DB			
QY	1921	ATTCAACAAACATGTCATCGTTTCAAAATTTTATTTTCAATGGCGTTATTTGTTTGTGT	1980
DB			
QY	1921	ATTCAACAAACATGTCATCGTTTCAAAATTTTATTTTCAATGGCGTTATTTGTTTGTGT	1980
DB			
QY	1981	AAATATCTGTTTAACTGCTGACGAATTTTTTAAATTTTTCAAAGAGAACATTTTGTAT	2040
DB			
QY	1981	AAATATCTGTTTAACTGCTGACGAATTTTTTAAATTTTTCAAAGAGAACATTTTGTAT	2040
DB			
QY	2041	ATAAATAACATTTTATGGAACACCGCTTAAGCTCGATGATTTTGGTTTTAGTTTTGTGC	2100
DB			
QY	2041	ATAAATAACATTTTATGGAACACCGCTTAAGCTCGATGATTTTGGTTTTAGTTTTGTGC	2100
DB			
QY	2101	GTTTGTGAAATCATTTAAACGACCTACATTTGATCCCTCACTTCTTTAATAATTAGGAATC	2160
DB			

Db 2101 GTTTTGTGAATCATTAACGACCTACATTTGATCCCTCATTTAAATTAATAGCAATC 2160
QY 2161 AAACATGATGATTAAAGTTTACCAAGAGACGTCTCTTATGGCTATTAAAGAGTCAGACGCAAG 2220
Db 2161 AAACATGATGATTAAAGTTTACCAAGAGACGTCTCTTATGGCTATTAAAGAGTCAGACGCAAG 2220
QY 2221 GATGACCGGGGTCAATTAAGACGCTCTTATATCAACCATTTACTCCACTAATTTGCTAATTA 2280
Db 2221 GATGACCGGGGTCAATTAAGACGCTCTTATATCAACCATTTACTCCACTAATTTGCTAATTA 2280
QY 2281 TCAGATTAAATTTGTTTAAATACGATTAATGTTTGAATTAAGTAGCTCTCAGCCACAGGC 2340
Db 2281 TCAGATTAAATTTGTTTAAATACGATTAATGTTTGAATTAAGTAGCTCTCAGCCACAGGC 2340
QY 2341 AAAGGATAAAATTTGGATTATCAAGAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
Db 2341 AAAGGATAAAATTTGGATTATCAAGAGATTTGTGGGCTTCCAAAAGCATAGGGAATGGCT 2400
QY 2401 CCAACTACATTTGGGAAATATATTTAAATCTCCATTTCCCATTTGCCACAGTCGTTGGA 2460
Db 2401 CCAACTACATTTGGGAAATATATTTAAATCTCCATTTCCCATTTGCCACAGTCGTTGGA 2460
QY 2461 GTTTATTTTCTTCCAAAGTGGAGAAATCAATTAATTTGCGAAATTTCTAATTCCTA 2520
Db 2461 GTTTATTTTCTTCCAAAGTGGAGAAATCAATTAATTTGCGAAATTTCTAATTCCTA 2520
QY 2521 CTGTGTGAAACCAAAACCAAGCAAAATTTAGTTAGAAATGTACGGAATCTACTATAGA 2580
Db 2521 CTGTGTGAAACCAAAACCAAGCAAAATTTAGTTAGAAATGTACGGAATCTACTATAGA 2580
QY 2581 ATTATATPAAACATATCAATATAGCTTTGACTTTTAAATTTAAACAAATTTATGTGGCAA 2640
Db 2581 ATTATATPAAACATATCAATATAGCTTTGACTTTTAAATTTAAACAAATTTATGTGGCAA 2640
QY 2641 TTAGTTAGATATTTTACCAGGGGAATTAAGCAGCACTGTTGATGCACTCTCTTCTTAAT 2700
Db 2641 TTAGTTAGATATTTTACCAGGGGAATTAAGCAGCACTGTTGATGCACTCTCTTCTTAAT 2700
QY 2701 ACTATACCTTTGAAAGAACTTATATGCTACTGTATTGCCAGTTGCCACTAATATATAACA 2760
Db 2701 ACTATACCTTTGAAAGAACTTATATGCTACTGTATTGCCAGTTGCCACTAATATATAACA 2760
QY 2761 ACATTTCACTTTGTGACATCGCTGTAATGAAGTTTGGAAACGACCTCTTAAACACTAAT 2820
Db 2761 ACATTTCACTTTGTGACATCGCTGTAATGAAGTTTGGAAACGACCTCTTAAACACTAAT 2820
QY 2821 AGGGTTAATTAACCACTAAATTTTCCAAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
Db 2821 AGGGTTAATTAACCACTAAATTTTCCAAAGTTGTCATTTTGTCTTAATGTGAGACGTAAT 2880
QY 2881 ATCTAATACGTCGGTCTAACCTAGAGTTGGTTCCGATCACAAATTTTGTGAGAGTACCT 2940
Db 2881 ATCTAATACGTCGGTCTAACCTAGAGTTGGTTCCGATCACAAATTTTGTGAGAGTACCT 2940
QY 2941 TTCATAAAAAATTTGTTGTTATATTTCAACGCTTTAAGTAAACCTTAATTAAGAGCT 3000
Db 2941 TTCATAAAAAATTTGTTGTTATATTTCAACGCTTTAAGTAAACCTTAATTAAGAGCT 3000
QY 3001 ATTTTCTATTTGATGTGAATTTGAAATGTCTTATATAAATAGTATGGAAGGGAATG 3060
Db 3001 ATTTTCTATTTGATGTGAATTTGAAATGTCTTATATAAATAGTATGGAAGGGAATG 3060
QY 3061 TAAATTAATATAGACCACAGATACAAAAGATGTCCCGTCTTAAACGCTCTGAGTCATTG 3120
Db 3061 TAAATTAATATAGACCACAGATACAAAAGATGTCCCGTCTTAAACGCTCTGAGTCATTG 3120
QY 3121 TCGTACCCTTTTGCCAACTTTTCAAGTTTCTTTCGTTGAAATGACTACACTTTTAAAA 3180
Db 3121 TCGTACCCTTTTGCCAACTTTTCAAGTTTCTTTCGTTGAAATGACTACACTTTTAAAA 3180
QY 3181 TAAATTGACAGATGTTGTTGATGATATTAATATTCGCAAAATTCACCCCTT 3240
Db 3181 TAAATTGACAGATGTTGTTGATGATATTAATATTCGCAAAATTCACCCCTT 3240

Db 3181 TAAATTGACAGATGTTGTTGATGATATTAATATTCGCAAAATTCACCCCTT 3240
QY 3241 AACCAATAATGGGTAAATGGGATATAAAATAGTTTAAACATAAAACAAAAATTTGCAAAATTTTG 3300
Db 3241 AACCAATAATGGGTAAATGGGATATAAAATAGTTTAAACATAAAACAAAAATTTGCAAAATTTTG 3300
QY 3301 AAAAATAGAGAGTAAATTTGATTTTTTTTAAAGAGTTTGAATTTGAAGTGAAGAAAAATATA 3360
Db 3301 AAAAATAGAGAGTAAATTTGATTTTTTTTAAAGAGTTTGAATTTGAAGTGAAGAAAAATATA 3360
QY 3361 ATAAAAATTAATAACCTGTAGTTTGATATATATAGTTTAGATAGCTCAAGTTTGTAGTAAT 3420
Db 3361 ATAAAAATTAATAACCTGTAGTTTGATATATATAGTTTAGATAGCTCAAGTTTGTAGTAAT 3420
QY 3421 TGAAGTCTTTGAATTTACTTTATATGTTTCTCACAGATATATATTTTCTGTTCTATCCC 3480
Db 3421 TGAAGTCTTTGAATTTACTTTATATGTTTCTCACAGATATATATTTTCTGTTCTATCCC 3480
QY 3481 AAGAATTCGGATAATTTCTCTATATTCGAGGCTCTCTCTTAAAGAGTCCGTTGATAAT 3540
Db 3481 AAGAATTCGGATAATTTCTCTATATTCGAGGCTCTCTCTTAAAGAGTCCGTTGATAAT 3540
QY 3541 CTTAAGCCCTTACTTGACACAGGCTCTATTTAAAGCCCAATTAATTTTCTTTTC 3600
Db 3541 CTTAAGCCCTTACTTGACACAGGCTCTATTTAAAGCCCAATTAATTTTCTTTTC 3600
QY 3601 AAAGCCCAACAGCTCAAGAGGAGAAAGTGGTTCGGTTTGGATTGAAAAAGTGGCG 3660
Db 3601 AAAGCCCAACAGCTCAAGAGGAGAAAGTGGTTCGGTTTGGATTGAAAAAGTGGCG 3660
QY 3661 GGCGGTTGGAACCTGAATCTTAAACCTTCACTCAATTCATCTTCCCGTCAAAATCTCT 3720
Db 3661 GGCGGTTGGAACCTGAATCTTAAACCTTCACTCAATTCATCTTCCCGTCAAAATCTCT 3720
QY 3721 GAAACTGAGCTTCGACGATGTTGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
Db 3721 GAAACTGAGCTTCGACGATGTTGGTTTACGAAGATCTGCGACGACCTTGTTCGACATCA 3780
QY 3781 GCCAGTCTCTGTTTACGTTAGTTCGGTTAGTCTGTCTAACACTTCAGCGCCACATTT 3840
Db 3781 GCCAGTCTCTGTTTACGTTAGTTCGGTTAGTCTGTCTAACACTTCAGCGCCACATTT 3840
QY 3841 GGGTTTTACGATAGGGTTTCAATAGATCGAATCGAATCATCTCTCTTGGTATTAGATTT 3900
Db 3841 GGGTTTTACGATAGGGTTTCAATAGATCGAATCGAATCATCTCTCTTGGTATTAGATTT 3900
QY 3901 CATCAAAATTCAAATCTATCGTCACCTAAAATCTCTCTTTCGGTATGTAGTATTTCCGG 3960
Db 3901 CATCAAAATTCAAATCTATCGTCACCTAAAATCTCTCTTTCGGTATGTAGTATTTCCGG 3960
QY 3961 GTTGTTCAAAAGTCGTCTTATTTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
Db 3961 GTTGTTCAAAAGTCGTCTTATTTGCTCTGTGATTTTGGCTTCTTACCCATTTTCAACAGTG 4020
QY 4021 CTATGTAGAGAAAGAACAAATCTTTGAAATCGAAAGTCTAAATGTATGTATCAATGTCT 4080
Db 4021 CTATGTAGAGAAAGAACAAATCTTTGAAATCGAAAGTCTAAATGTATGTATCAATGTCT 4080
QY 4081 ACATTATGAGATTCGCATGATTTATAGTCAACAGATCTCTTAAAGCGTTTACTTTTGT 4140
Db 4081 ACATTATGAGATTCGCATGATTTATAGTCAACAGATCTCTTAAAGCGTTTACTTTTGT 4140
QY 4141 GGAGCATTTCTTTGTAGTTCTATCAATAAAGTTCTAGGAATGATTAATCTTAAAGGAAG 4200
Db 4141 GGAGCATTTCTTTGTAGTTCTATGCAATAAAGTTCTAGGAATGATTAATCTTAAAGGAAG 4200
QY 4201 CATCTCAAAATGTGGCTAGTTCTTGTCTCAGGTTTAAACCAATTTTGCATTTTGCCTTTT 4260
Db 4201 CATCTCAAAATGTGGCTAGTTCTTGTCTCAGGTTTAAACCAATTTTGCATTTTGCCTTTT 4260
QY 4261 AGTTAGAAATTTGTGACTTGTCTAGCTTTTGTGACTTAACTCTGTCTCTGTGAGCAAAAGTTT 4320
Db 4261 AGTTAGAAATTTGTGACTTGTCTAGCTTTTGTGACTTAACTCTGTCTCTGTGAGCAAAAGTTT 4320

```
QY 4321 GATCAAAACCATTAACCTTATTTGATTTCTCTCTTAGATATATACATCAATTTATGATTTT 4380
Db |||||
QY 4321 GATCAAAACCATTAACCTTATTTGATTTCTCTCTTAGATATATACATCAATTTATGATTTT 4380
Db |||||
QY 4381 TCCTTCTCTACAGTTTCATCGGTTACAGTCCA 4413
Db |||||
QY 4381 TCCTTCTCTACAGTTTCATCGGTTACAGTCCA 4413
Db |||||

RESULT 2
AAF25319
ID AAF25319 standard; DNA; 4309 BP.
XX
AC AAF25319;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plant promoter insert in plasmid pBin19.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
OS Arabidopsis thaliana.
OS Unidentified.
XX
XX Key Location/Qualifiers
FH promoter 30..2178
FT /*tag= a
FT misc_feature 2179..4309
FT /*tag= b
FT /note= "GUS coding region"
XX
XX WO200100833-A1.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-FR001768.
XX
XX 25-JUN-1999; 99FR-00008185.
XX
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Hoffmann B, Mollier P, Pelletier G;
XX
XX WPI; 2001-102893/11.
XX
XX New constitutive plant promoter active specifically in roots, useful for
XX controlling expression of pest or drought resistance genes, and related
XX transgenic plants.
XX
XX Example 3; Page 73-75; 92pp; French.
XX
XX The present sequence represents a plant promoter insert in plasmid
XX pBin19. The plant promoter directs expression of a selected sequence in
XX root cells at all stages of development of a plant. The plant promoter is
XX used to control expression of genes in a root-specific manner, especially
XX genes that provide resistance to parasites, pests (nematodes or fungi),
XX water and salt stress, or alter sugar content or nitrogen transport.
XX Fragments of the promoter are useful as probes or primers to detect or
XX amplify at least part of the promoter
XX
XX Sequence 4309 BP; 1262 A; 868 C; 915 G; 1264 T; 0 U; 0 Other;
XX
XX Query Match 48.8%; Score 2154; DB 4; Length 4309;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 131 CTCAGTCGAATTTGTGATATATTGTAAGCAATCTGAAAAGATAAGTGGGATATAAAC 190
XX |||||
XX 25 CTCAGTCGAATTTGTGATATATTGTAAGCAATCTGAAAAGATAAGTGGGATATAAAC 84
```

```
QY 191 AACCGCGGAAAGTACAAAGTTCTACCTTTTGGCATGGAACCATGTTTTAGGATTTAC 250
Db |||||
QY 85 AACCGCGGAAAGTACAAAGTTCTACCTTTTGGCATGGAACCATGTTTTAGGATTTAC 144
Db |||||
QY 251 TTGTGAATTCCTCGAATCTTTTCATTTCTTGAATTTGATATTACATTTTATCAAAAAAAA 310
Db |||||
QY 145 TTGTGAATTCCTCGAATCTTTTCATTTCTTGAATTTGATATTACATTTTATCAAAAAAAA 204
Db |||||
QY 311 GTACAAAGTTCTACCAAGCACAGGAGTTAAACAACCTTGTGTCAATGCTAAATTTAAAG 370
Db |||||
QY 205 GTACAAAGTTCTACCAAGCACAGGAGTTAAACAACCTTGTGTCAATGCTAAATTTAAAG 264
Db |||||
QY 371 CCTAATCTTATGATTTTCCCTTTCTTCAAGATATATCTGATATGATATGACCCATTT 430
Db |||||
QY 265 CCTAATCTTATGATTTTCCCTTTCTTCAAGATATATCTGATATGATATGACCCATTT 324
Db |||||
QY 431 GTTTGTCAATTAACCTTCCACTCTATACATCAGTATCTCAAGTCGAATAACAATATCCAT 490
Db |||||
QY 325 GTTTGTCAATTAACCTTCCACTCTATACATCAGTATCTCAAGTCGAATAACAATATCCAT 384
Db |||||
QY 491 AAGAAGTGGTATATTGTGAAAAAAGTGGTATATCTGGTATATACAATACCA 550
Db |||||
QY 385 AAGAAGTGGTATATTGTGAAAAAAGTGGTATATCTGGTATATACAATACCA 444
Db |||||
QY 551 CGGTCTCGAATTCCTCAACAATTTCTAGGAGAAATGGAGCTGTCTCTTTGGTTTATT 610
Db |||||
QY 445 CGGTCTCGAATTCCTCAACAATTTCTAGGAGAAATGGAGCTGTCTCTTTGGTTTATT 504
Db |||||
QY 611 TTATTTCTTAATAACATACCTCTATATTTAAACAATTCGATGTCTCGTTAAATTTTCAAT 670
Db |||||
QY 505 TTATTTCTTAATAACATACCTCTATATTTAAACAATTCGATGTCTCGTTAAATTTTCAAT 564
Db |||||
QY 671 GTCCCTAAATTTCTTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATG 730
Db |||||
QY 565 GTCCCTAAATTTCTTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATG 624
Db |||||
QY 731 CATAGGGCAGTAGTTTACCTTTTAAACCATCAAAATATATTAATAGAAAGGAACCTTC 790
Db |||||
QY 625 CATAGGGCAGTAGTTTACCTTTTAAACCATCAAAATATATTAATAGAAAGGAACCTTC 684
Db |||||
QY 791 CTAAAAGAACAAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGACAGAGAAACG 850
Db |||||
QY 685 CTAAAAGAACAAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGACAGAGAAACG 744
Db |||||
QY 851 TATGGCGGACCTCGTAAACAGGAGCGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 910
Db |||||
QY 745 TATGGCGGACCTCGTAAACAGGAGCGTCCCGACCACTGCGGAGACGGCGAGACGCTGAC 804
Db |||||
QY 911 TGATTTTCTTTTCTTTTCTTTTCTTAAAGACGTTTGTTCGTGCTTACAGGCTCAAAACCA 970
Db |||||
QY 805 TGATTTTCTTTTCTTTTCTTTTCTTAAAGACGTTTGTTCGTGCTTACAGGCTCAAAACCA 864
Db |||||
QY 971 TATCCAAATGTTCTGCTATTTATATATAACTAAAGATCCCTCTTGTGCTTTGCTTTA 1030
Db |||||
QY 865 TATCCAAATGTTCTGCTATTTATATATAACTAAAGATCCCTCTTGTGCTTTGCTTTA 924
Db |||||
QY 1031 TTCTGATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1090
Db |||||
QY 925 TTCTGATATATATATATATATATATATATATATATATATATATATATATATATATATAT 984
Db |||||
QY 1091 TACTGACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAATGACGCCCAAAATTTG 1150
Db |||||
QY 985 TACTGACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAATGACGCCCAAAATTTG 1044
Db |||||
QY 1151 AAGAGTTTCTTTCTTCTGCAACTCTTATCTTACTCATGAGCTATGTTAATATGTA 1210
Db |||||
QY 1045 AAGAGTTTCTTTCTTCTGCAACTCTTATCTTACTCATGAGCTATGTTAATATGTA 1104
Db |||||
QY 1211 ATGTTGGCACTCTCGTATTTAAATATGCACTTGCACCTAGATATAAAGATGATAGACA 1270
Db |||||
QY 1105 ATGTTGGCACTCTCGTATTTAAATATGCACTTGCACCTAGATATAAAGATGATAGACA 1164
Db |||||
```

```
Qy 1271 TTTAGTTTAAACCTGAAATGTTATTTGAACCTCTTTGGATTACGTGGATTCTTGTATGGA 1330
Db 1165 TTTAGTTTAAACCTGAAATGTTATTTGAACCTCTTTGGATTACGTGGATTCTTGTATGGA 1224
Qy 1331 TTTAAATTTTGAAGATATTTATATATTTGAAGATGTTTATATATATAGAGTTTATATAGCA 1390
Db 1225 TTTAAATTTTGAAGATATTTATATATTTGAAGATGTTTATATATATAGAGTTTATATAGCA 1284
Qy 1391 GAAATATTTGATAGATGTTGCTCTTTTGTAGTTACTCTTTTGTGCGTAGTCCCTTT 1450
Db 1285 GAAATATTTGATAGATGTTGCTCTTTTGTAGTTACTCTTTTGTGCGTAGTCCCTTT 1344
Qy 1451 CTCCTATPCCCTCATGAAAGAAAATCCAAATAGTTTAAAGAAAATTTTGTGTAAATTCAT 1510
Db 1345 CTCCTCATCTCTCATGAAAGAAAATCCAAATAGTTTAAAGAAAATTTTGTGTAAATTCAT 1404
Qy 1511 AGTCTTTTTCGTAAACACAGTCTATGTAGCTATCGTCATCATATATCTCTTTCGCAACA 1570
Db 1405 AGTCTTTTTCGTAAACACAGTCTATGTAGCTATCGTCATCATATATCTCTTTCGCAACA 1464
Qy 1571 CAAAAAAGATCGTTTGTAAATTTAGTAGGGCACTAAAGTCGTCATTGTGTCTCTGT 1630
Db 1465 CAAAAAGATCGTTTGTAAATTTAGTAGGGCACTAAAGTCGTCATTGTGTCTCTGT 1524
Qy 1631 CGAAATCTAGCGTCTGTATCCCAATAAAGTTGTGTGATTCGAGCTTCCAAAGATTATA 1690
Db 1525 CGAAATCTAGCGTCTGTATCCCAATAAAGTTGTGTGATTCGAGCTTCCAAAGATTATA 1584
Qy 1691 ATCTTTTTCAGTGGTCAATGAAGTTCTAACTTCGATACGAGTGTATCCATATTAAT 1750
Db 1585 ATCTTTTTCAGTGGTCAATGAAGTTCTAACTTCGATACGAGTGTATCCATATTAAT 1644
Qy 1751 TCTAACATATACGCTCTGTTTTTGGTAGGCTCTCGTCTTTTGAGACACACCCCTTGCTA 1810
Db 1645 TCTAACATATACGCTCTGTTTTTGGTAGGCTCTCGTCTTTTGAGACACACCCCTTGCTA 1704
Qy 1811 ATGTTTGTTCGACCTTAGACAATCCATAATACGTTAGCGTAGTCGAAGTTGCGACCAAAA 1870
Db 1705 ATGTTTGTTCGACCTTAGACAATCCATAATACGTTAGCGTAGTCGAAGTTGCGACCAAAA 1764
Qy 1871 TGGTCCAAATATAATTTAAATTTGGCCCAACAAACAACTTTTACAAACAATTTCAACAAA 1930
Db 1765 TGGTCCAAATATAATTTAAATTTGGCCCAACAAACAACTTTTACAAACAATTTCAACAAA 1824
Qy 1931 CATGCATCGTTTCAAAATTTTATTTTCAATGGCGTTATTTGTTCATTGTAAATATCTG 1990
Db 1825 CATGCATCGTTTCAAAATTTTATTTTCAATGGCGTTATTTGTTCATTGTAAATATCTG 1884
Qy 1991 TTAACTCAGTCGAGAAATTTTAAATTTTCAAGAGAAACAATTTTGGATATAAATAACA 2050
Db 1885 TTAACTCAGTCGAGAAATTTTAAATTTTCAAGAGAAACAATTTTGGATATAAATAACA 1944
Qy 2051 TTTTATGGAACCCCGGTTAGCTCGATGATTTTGGAGTTTGTAGTTTGTCTGTTTGTGAA 2110
Db 1945 TTTTATGGAACCCCGGTTAGCTCGATGATTTTGGAGTTTGTAGTTTGTCTGTTTGTGAA 2004
Qy 2111 ATCAATTAACGACCTACATTTGATCCCTCATTTACTTCTTAAATATAGGAATCAACATGATG 2170
Db 2005 ATCAATTAACGACCTACATTTGATCCCTCATTTACTTCTTAAATATAGGAATCAACATGATG 2064
Qy 2171 ATTAAGTTTCAACAAAGAGCTCTTATAGGCTTATTAAGAGTCAGCGCAAGGATGACCGGG 2230
Db 2065 ATTAAGTTTCAACAAAGAGCTCTTATAGGCTTATTAAGAGTCAGCGCAAGGATGACCGGG 2124
Qy 2231 GTCAATTAAGAGCTCTTATTAATTAACCACTTACTCCACTAAATTTGCTAAATTAATCAG 2284
Db 2125 GTCAATTAAGAGCTCTTATTAATTAACCACTTACTCCACTAAATTTGCTAAATTAATCAG 2178
```

RESULT 3

AAF25316

ID AAF25316 standard; DNA; 2149 BP.

XX

```
AC AAF25316;
XX 30-APR-2001 (first entry)
XX Nucleotide sequence of a plant promoter for expression in roots.
DE Plant promoter; root cell; root-specific expression; parasite resistance;
XX nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
OS Arabidopsis thaliana.
XX WO2001:00833-A1.
XX 04-JAN-2001.
XX 23-JUN-2000; 2000WO-FR001768.
XX 25-JUN-1999; 99FR-00008185.
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX Hoffmann B, Mollier P, Pelletier G;
XX WPI; 2001-102893/11.
XX New constitutive plant promoter active specifically in roots, useful for
XX controlling expression of pest or drought resistance genes, and related
XX transgenic plants.
XX Claim 1; Fig 1; 92pp; French.
XX The present sequence represents a plant promoter that directs expression
XX of a selected sequence in root cells at all stages of development of a
XX plant. The plant promoter is used to control expression of genes in a
XX root-specific manner, especially genes that provide resistance to
XX parasites, pests (nematodes or fungi), water and salt stress, or alter
XX sugar content or nitrogen transport. Fragments of the promoter are useful
XX as probes or primers to detect or amplify at least part of the promoter
XX
SQ Sequence 2149 BP; 693 A; 365 C; 335 G; 756 T; 0 U; 0 Other;
Query Match 48.7%; Score 2149; DB 4; Length 2149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 136 GTCCGAATTTGATATATTTGAAGCAATCTGAAAGAAATAGTGGATATATAAACAACCG 195
Db 1 GTCCGAATTTGATATATTTGAAGCAATCTGAAAGAAATAGTGGATATATAAACAACCG 60
Qy 196 GCGAAGGTACAAGTCTACCTTTTGGCATGGAACCATGTTTGGATTACTTTGT 255
Db 61 GCGAAGGTACAAGTCTACCTTTTGGCATGGAACCATGTTTGGATTACTTTGT 120
Qy 256 AATTCTCTGAATCTTTCATTTCTGAATGATATTTATCATTTTATCAAAAAAAGATACA 315
Db 121 AATTCTCTGAATCTTTCATTTCTGAATGATATTTATCATTTTATCAAAAAAAGATACA 180
Qy 316 AGTTCTACCAAGCACAGGAGTTAAACAACCTGTGTCTCAAACTGTAATTTAAAGCCTAA 375
Db 181 AGTTCTACCAAGCACAGGAGTTAAACAACCTGTGTCTCAAACTGTAATTTAAAGCCTAA 240
Qy 376 TCTTATGATTTCCCTTTTCTTCCAGTATATCTGATATTGATATGACCCATTGTTTG 435
Db 241 TCTTATGATTTCCCTTTTCTTCCAGTATATCTGATATTGATATGACCCATTGTTTG 300
Qy 436 TCATTAACTTCCCACTCTATACATCAGTATCTCAAGTCGAATAACAATATCCATAAGAA 495
Db 301 TCATTAACTTCCCACTCTATACATCAGTATCTCAAGTCGAATAACAATATCCATAAGAA 360
Qy 496 GTGGTATATTGTGAAAAAAGTGTATGTTACTGTTATATACATACCAACCGTC 555
Db 361 GTGGTATATTGTGAAAAAAGTGTATGTTACTGTTATATACATACCAACCGTC 420
```

QY 556 TCGAATTCCTCAACAAATTCCTAGGAGAAAATGAGCGTGTCTCTTTGGTTTATTTTATT 615
Db 421 TCGAATTCCTCAACAAATTCCTAGGAGAAAATGAGCGTGTCTCTTTGGTTTATTTTATT 480
QY 616 CTTTAATAACATCTCTATATTTTAAACACATTCGATGCTCGCTTAAATTTTCAATGTC 675
Db 481 CTTTAATAACATCTCTATATTTTAAACATTCGATGCTCGCTTAAATTTTCAATGTC 540
QY 676 TAAATTTCTCTAATCATPAATTCGTAAGAAAAATTCGTCGAAGCCACAGGACATGTCATAG 735
Db 541 TAAATTTCTCTAATCATPAATTCGTAAGAAAAATTCGTCGAAGCCACAGGACATGTCATAG 600
QY 736 GGCACGTAGTACTCTTTTAAACCATCAAAAATATATTAATAGAAAAGGAACCTTCTTAAA 795
Db 601 GGCACGTAGTACTCTTTTAAACCATCAAAAATATATTAATAGAAAAGGAACCTTCTTAAA 660
QY 796 AGAACCAATTTAATAAGTGGATAAAAAAGATAAGAAAGTAGGAGAGAAAAACGTATGG 855
Db 661 AGAACCAATTTAATAAGTGGATAAAAAAGATAAGAAAGTAGGAGAGAAAAACGTATGG 720
QY 856 CCGGAGCTGTAAACAGGAGCGTCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 915
Db 721 CCGGAGCTGTAAACAGGAGCGTCCGACCACTGCGGAGACGGCGAGACGCTGACTGATT 780
QY 916 TTTTCTTTCTTTTCTTAAGAACGTTGTTTCGTCTTACAAGGTCAAAACCATATCC 975
Db 781 TTTTCTTTCTTTTCTTAAGAACGTTGTTTCGTCTTACAAGGTCAAAACCATATCC 840
QY 976 AATTGTTCTGCCCTATTATATATAAATACTAAAAGATCCCCCTGTCGCTTTGCTTATTTCGT 1035
Db 841 AATTGTTCTGCCCTATTATATATAAATACTAAAAGATCCCCCTGTCGCTTTGCTTATTTCGT 900
QY 1036 GATATATAATCTAACTAAATAGTCTTAAAAATATATATATGTCCTACTGTTTCTACTG 1095
Db 901 GATATATAATCTAACTAAATAGTCTTAAAAATATATATATGTCCTACTGTTTCTACTG 960
QY 1096 ACCTCAGTCCCTAGTGTAGCTATGACATATGTGAAAATGAGCCCAAAATTTGAAG 1155
Db 961 ACCTCAGTCCCTAGTGTAGCTATGACATATGTGAAAATGAGCCCAAAATTTGAAG 1020
QY 1156 TTCTCTCTCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1215
Db 1021 TTCTCTCTCTGCAACTAACTCTTATCTTACTCATTGAGCTATGTTAAATATTGAATGTT 1080
QY 1216 GGCACCTCTGATTAATATGCGAGTTGCACTAGATAAAAAACATGATAGCATTTAG 1275
Db 1081 GGCACCTCTGATTAATATGCGAGTTGCACTAGATAAAAAACATGATAGCATTTAG 1140
QY 1276 TTTTAAACTTGAATGTTATTTGCAACTCTTTGGATTACGTGGATTGTTGATGATTAAA 1335
Db 1141 TTTTAAACTTGAATGTTATTTGCAACTCTTTGGATTACGTGGATTGTTGATGATTAAA 1200
QY 1336 TTTTGAAGATTTATATATGAAAGATGTTTATATATATATATATATATATATAGCAGAAAA 1395
Db 1201 TTTTGAAGATTTATATATGAAAGATGTTTATATATATATATATATATATAGCAGAAAA 1260
QY 1396 TATTGATGATAGTGTGTCCTTTTGTAGTACTCTTTTGTGCGTAGTCCCTTCTCCT 1455
Db 1261 TATTGATGATAGTGTGTCCTTTTGTAGTACTCTTTTGTGCGTAGTCCCTTCTCCT 1320
QY 1456 CATCTCTCTATGAAGAAAAATCAAAATAGTTTAAAGGAAATTTTGTGTAATTCATAGTCT 1515
Db 1321 CATCTCTCTATGAAGAAAAATCAAAATAGTTTAAAGGAAATTTTGTGTAATTCATAGTCT 1380
QY 1516 TTTTCGTAAACCAAGTCTTAGTATGCTCATCATATCTCTTTTGCACAAACAAAA 1575
Db 1381 TTTTCGTAAACCAAGTCTTAGTATGCTCATCATATCTCTTTTGCACAAACAAAA 1440
QY 1576 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAAGTCGTCATTGTTGTCCTGTCGAAA 1635
Db 1441 AAGATCGTTTTTGTAAAAATTTAGTAGGCACTAAAGTCGTCATTGTTGTCCTGTCGAAA 1500

QY 1636 TCTAGCGTTCTGTCATCCACAAATAGTTCGTTTGGTTTCGAGCTTCCAGATTATAATCTT 1695
Db 1501 TCTAGCGTTCTGTCATCCACAAATAGTTCGTTTGGTTTCGAGCTTCCAGATTATAATCTT 1560
QY 1696 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTTATACGAGTGTATCCATATAATTTCTAA 1755
Db 1561 TTTTAGATGGGTCAATGAAGATTTCTAACTTCGTTATACGAGTGTATCCATATAATTTCTAA 1620
QY 1756 CATATACGTCTCTGTTTGGTAGGCTCTGCGTCTTTTGAGACCAACCCCTTGTCTAAATGTT 1815
Db 1621 CATATACGTCTCTGTTTGGTAGGCTCTGCGTCTTTTGAGACCAACCCCTTGTCTAAATGTT 1680
QY 1816 TTCTTCACCTTAGAGCAATCCATAATACGTTTACGTGAGTCGAAAGTTCACCAAAATCGTC 1875
Db 1681 TTCTTCACCTTAGAGCAATCCATAATACGTTTACGTGAGTCGAAAGTTCACCAAAATCGTC 1740
QY 1876 CAAATATAATTTAAATTTGGCCCAAAAAACAACATTTTACAAAACAAATTTCAACAAACATGC 1935
Db 1741 CAAATATAATTTAAATTTGGCCCAAAAAACAACATTTTACAAAACAAATTTCAACAAACATGC 1800
QY 1936 ATCGTTTCAAAATTTTATTTATTCATAGGCGTTATTTGTCATTTGTAATATTTCTGTTTAA 1995
Db 1801 ATCGTTTCAAAATTTTATTTATTCATAGGCGTTATTTGTCATTTGTAATATTTCTGTTTAA 1860
QY 1996 CTCACGTACGAATTTTAAATTTTCAAGAGAACATTTTTCATATAAAATAACATTTTA 2055
Db 1861 CTCACGTACGAATTTTAAATTTTCAAGAGAACATTTTTCATATAAAATAACATTTTA 1920
QY 2056 TGAACACACCGGTAAAGCTCGATGATTTTGAGTTTGTAGTTTTCGTTTGTGAATCAT 2115
Db 1921 TGAACACACCGGTAAAGCTCGATGATTTTGAGTTTGTAGTTTTCGTTTGTGAATCAT 1980
QY 2116 TAAGACGTACATTTGATCCCTCATTTACTTAAATATAGGAATCAAAAGATGATTA 2175
Db 1981 TAAGACGTACATTTGATCCCTCATTTACTTAAATATAGGAATCAAAAGATGATTA 2040
QY 2176 GTTCACCAAGAGCTCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATGACCGGGTCAT 2235
Db 2041 GTTCACCAAGAGCTCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATGACCGGGTCAT 2100
QY 2236 TAAGACGTCTTATATTCAACCATTTACTCCACTAATTTGCTAATTAATCAG 2284
Db 2101 TAAGACGTCTTATATTCAACCATTTACTCCACTAATTTGCTAATTAATCAG 2149

RESULT 4
AAF25317
ID AAF25317 standard; DNA; 4280 BP.
XX
AC AAF25317;
XX
XX
DT 30-APR-2001 (first entry)
XX
XX
DE Nucleotide sequence of a root promoter and the GSU gene coding region.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; GUS gene; ss.
XX
OS Synthetic.
OS Arabidopsis thaliana.
OS Unidentified.
XX
XX WO200100833-A1.
XX
PD 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-FR001768.
XX
XX 25-JUN-1999; 99FR-00008185.
PR
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA
XX

PI Hoffmann B, Mollier P, Pelletier G;
XX WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful for
PT controlling expression of pest or drought resistance genes, and related
PT transgenic plants.
XX
PS Claim 5; Page 69-71; 92pp; French.
XX
XX The present sequence represents a plant promoter linked to a GUS coding
CC region. The plant promoter directs expression of a selected sequence in
CC root cells at all stages of development of a plant. The plant promoter is
CC used to control expression of genes in a root-specific manner, especially
CC genes that provide resistance to parasites, pests (nematodes or fungi),
CC water and salt stress, or alter sugar content or nitrogen transport.
CC Fragments of the promoter are useful as probes or primers to detect or
CC amplify at least part of the promoter
XX
SQ Sequence 4280 BP; 1256 A; 859 C; 907 G; 1258 T; 0 U; 0 Other;
Query Match 48.7%; Score 2149; DB 4; Length 4280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 GTGCAATTGTGATATATTGTAAGCAATCTGAAAGAAATGAGTGGGATATATAAACAACCG 195
DB 1 GTGCAATTGTGATATATTGTAAGCAATCTGAAAGAAATGAGTGGGATATATAAACAACCG 60
QY 196 GCGAAGTACAGTTCTACCTTTTGGCATGGAACCATGTTTATAGATTACTTTGT 255
DB 61 GCGAAGTACAGTTCTACCTTTTGGCATGGAACCATGTTTATAGATTACTTTGT 120
QY 256 AATTCCTGAACTCTTCAATTCCTGAATTCGATATTTACATTTTATCAAAAAAAGTACA 315
DB 121 AATTCCTGAACTCTTCAATTCCTGAATTCGATATTTACATTTTATCAAAAAAAGTACA 180
QY 316 AGTTCTACAAAGCAGAGTTAAACAACCTTGTGTGCAAAATGCTAATTTAAAGCCTAA 375
DB 181 AGTTCTACAAAGCAGAGTTAAACAACCTTGTGTGCAAAATGCTAATTTAAAGCCTAA 240
QY 376 TCTTATGATTTCCCTTTCTTCCACATATATATGATATGATATGATACCCATTTGTTG 435
DB 241 TCTTATGATTTCCCTTTCTTCCACATATATATGATATGATATGATACCCATTTGTTG 300
QY 436 TCATTAACCTTCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 495
DB 301 TCATTAACCTTCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATATCCATAAGAA 360
QY 496 GTGGTATATTGTGAAAAAAGTGGTATACGTGATATACAAATACCAACCGGTC 555
DB 361 GTGGTATATTGTGAAAAAAGTGGTATACGTGATATACAAATACCAACCGGTC 420
QY 556 TCGAATTCGCTCAAAATTTCTAGAGAAAAAGTGGTGTCTCTTTGTTGTTTATTTAT 615
DB 421 TCGAATTCGCTCAAAATTTCTAGAGAAAAAGTGGTGTCTCTTTGTTGTTTATTTAT 480
QY 616 CTTAATAACATCTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTTGAATGTGCC 675
DB 481 CTTAATAACATCTCTATATTTTAAACACTTCGATGTCTCGCTTAAATTTTGAATGTGCC 540
QY 676 TAAATTTCTTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATCATAG 735
DB 541 TAAATTTCTTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATCATAG 600
QY 736 GGCAGTAGTTACCTTTTAAACCATCAAAATATATTAATAGAAAGGAACTTCTTAA 795
DB 601 GGCAGTAGTTACCTTTTAAACCATCAAAATATATTAATAGAAAGGAACTTCTTAA 660
QY 796 AGAACAAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGCAGAGAAAAACGTATGG 855
DB 661 AGAACAAATTTAATAAGTGGATAAAAAAGATAAGAGGTAGGCAGAGAAAAACGTATGG 720

QY 856 CCGGACTCGTAAACAAGGACGCTCCGACCACTGCGGAGACGCGGAGACGCTGACTGATT 915
DB 721 CCGGACTCGTAAACAAGGACGCTCCGACCACTGCGGAGACGCGGAGACGCTGACTGATT 780
QY 916 TTTTCTTTTCTTTTCTTAAAGAAAGTGTGTTTCTGCTTCTAAGGTCGAAACCATATCC 975
DB 781 TTTTCTTTTCTTTTCTTAAAGAAAGTGTGTTTCTGCTTCTAAGGTCGAAACCATATCC 840
QY 976 AATTCGTCTGCTTATTTATATAAAGATCCCTCTTCTGCTTCTTATTCGT 1035
DB 841 AATTCGTCTGCTTATTTATATAAAGATCCCTCTTCTGCTTCTTATTCGT 900
QY 1036 GATATATAATCTAACTTAAATTTAGTCTTAAATATATATGCTCTACCTATGTTTCTACTG 1095
DB 901 GATATATAATCTAACTTAAATTTAGTCTTAAATATATATGCTCTACCTATGTTTCTACTG 960
QY 1096 ACCTCAGTCCCTAGTGTAGCTATATGACATATGTGAATGACGCCCAAAATTTGAAGAG 1155
DB 961 ACCTCAGTCCCTAGTGTAGCTATATGACATATGTGAATGACGCCCAAAATTTGAAGAG 1020
QY 1156 TTCTCTTCTCCGCAACTAACTCTTATCTTACTCATTTGAGCTATGTTAAATATTGAATGTT 1215
DB 1021 TTCTCTTCTCCGCAACTAACTCTTATCTTACTTCTGAGCTATGTTAAATATTGAATGTT 1080
QY 1216 GGCACCTCTGATTTAAATATGCCAGTTGCACCTAGATAAAAAATCATGATAGACATTTAG 1275
DB 1081 GGCACCTCTGATTTAAATATGCCAGTTGCACCTAGATAAAAAATCATGATAGACATTTAG 1140
QY 1276 TTTAAACCTTGAAATGTTTATTTGAACTCTTTGGATTACGTGGATGTTGTTAGTAA 1335
DB 1141 TTTAAACCTTGAAATGTTTATTTGAACTCTTTGGATTACGTGGATGTTGTTAGTAA 1200
QY 1336 TTTTCAAGATATTTATATATTGAAGATGTTTATATATATTAGATTTATATAGAGAAAA 1395
DB 1201 TTTTGAAGATTTATATATTGAAGATGTTTATATATTAGATTTATATAGAGAAAA 1260
QY 1396 TATTGATGTAGATGTTGCTTTTGTAGTACTCTTTTGTGCTAGTCTCTTCTCTCT 1455
DB 1261 TATTGATGTAGATGTTGCTTTTGTAGTACTCTTTTGTGCTAGTCTCTTCTCTCT 1320
QY 1456 CATCTCCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1515
DB 1321 CATCTCCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCATAGTCT 1380
QY 1516 TTTTCTGTAACACAGTCTTATGTAGCTATCGTCATCATATTCTCTTTGCAACAACAAA 1575
DB 1381 TTTTCTGTAACACAGTCTTATGTAGCTATCGTCATCATATTCTCTTTGCAACAACAAA 1440
QY 1576 AAGATCGTTTTTGTAAAAATTTAGTAGGGGACATAAAGTCGTCAATTTGTTGCTGCGAAA 1635
DB 1441 AAGATCGTTTTTGTAAAAATTTAGTAGGGGACATAAAGTCGTCAATTTGTTGCTGCGAAA 1500
QY 1636 TCTAGCTTCTGTCTATCCACAAATTAAGTTGTTGATTCGAGCTTCCAGATTAATATCTT 1695
DB 1501 TCTAGCTTCTGTCTATCCACAAATTAAGTTGTTGATTCGAGCTTCCAGATTAATATCTT 1560
QY 1696 TTTTATAGTGGTCTAAGAAATTTCTAATCTCGTATACAGGTGTATCCATATAATTTCTAA 1755
DB 1561 TTTTATAGTGGTCTAAGAAATTTCTAATCTCGTATACAGGTGTATCCATATAATTTCTAA 1620
QY 1756 CATATACGTCTGTTTTTGTAGGCTCTGCGCTCTTTTGTAGACCAACCCCTTGTCTAAATGTT 1815
DB 1621 CATATACGTCTGTTTTTGTAGGCTCTGCGCTCTTTTGTAGACCAACCCCTTGTCTAAATGTT 1680
QY 1816 TTTTGTGACCTTAGACATCCATTAATAGCTTACGTGAGTCGAGTTGACCAAAATGTC 1875
DB 1681 TTTTGTGACCTTAGACATCCATTAATAGCTTACGTGAGTCGAGTTGACCAAAATGTC 1740
QY 1876 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTTCAACAACATGC 1935
DB 1741 CAAATATAATTTAAATTTGGCCACAAAACAACATTTTACAAACAAATTTCAACAACATGC 1800
QY 1936 ATCGTTTCAAAATTTATTTATTCAAATGGCGTTATTTGTTTCATTTGTAATATTTCTGTTTAA 1995

Query Match 0.6%; Score 26; DB 6; Length 15046;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3349 AAAAAATATATAATAAAAAATAATAA 3374
Db 9895 AAAAAATATATAATAAAAAATAATAA 9870

RESULT 7
ABV20030
ID ABV20030 standard; cDNA; 243 BP.
XX AC ABV20030;
XX DT 13-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 20021.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.
XX XX
XX PN WD200160860-A2.
XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.

XX PS Claim 1; Page 3266; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 243 BP; 126 A; 20 C; 39 G; 58 T; 0 U; 0 Other;

Query Match 0.6%; Score 25; DB 5; Length 243;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 TATATTGCGAAAAAATAAAAAA 524
Db 169 TATATTGCGAAAAAATAAAAAA 193

RESULT 8
ABV49789
ID ABV49789 standard; cDNA; 364 BP.
XX AC ABV49789;
XX DT 17-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 49780.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.

XX PS Claim 1; Page 9710; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 364 BP; 146 A; 72 C; 69 G; 76 T; 0 U; 1 Other;

Query Match 0.6%; Score 25; DB 5; Length 364;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 TATATTGCGAAAAAATAAAAAA 524
Db 207 TATATTGCGAAAAAATAAAAAA 231

RESULT 9
AAV40523
ID AAV40523 standard; cDNA; 2033 BP.

XX AC AAV40523;

XX DT 27-OCT-1998 (first entry)

XX XX

```
DE Homo sapiens CN483_2 clone secreted protein coding region.
XX secreted protein; CN483_2; ds.
KW Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 112..1215
FT /*tag= a
FT /*note= "secreted protein"
XX
PN WO9830695-A2.
XX
PD 16-JUL-1998.
XX
PF 09-JAN-1998; 98WO-US000543.
XX
PR 09-JAN-1997; 97US-00780814.
PR 08-JAN-1998; 98US-00004684.
XX
PA (GENY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
PI Spaulding V, Agostino MJ;
XX
DR WPI; 1998-413686/35.
DR P-PSDB; AAW29662.
XX
XX New isolated nucleic acids and secreted proteins - obtained from human
PT adult ovary, human foetal kidney, human foetal brain and human adult
PT brain cDNA libraries.
XX
PS Claim 22; Page 73-74; 113pp; English.
XX
CC The sequence is that encoding a novel, isolated secreted protein
XX
SQ Sequence 2033 BP; 611 A; 386 C; 517 G; 513 T; 0 U; 6 Other;
    Query Match 0.6%; Score 25; DB 2; Length 2033;
    Best Local Similarity 100.0%; Pred. No. 16;
    Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 TATATTGTGAAAAAAGAAAAA 524
Db 2009 TATATTGTGAAAAAAGAAAAA 2033

RESULT 10
ACH21196
ID ACH21196 standard; cDNA; 388 BP.
XX
AC ACH21196;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult liver cDNA #808.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I..
PA (STAC/) STACHE-CRAIN B.

PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 8408; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations,
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 388 BP; 115 A; 89 C; 95 G; 89 T; 0 U; 0 Other;
    Query Match 0.5%; Score 24; DB 8; Length 388;
    Best Local Similarity 100.0%; Pred. No. 46;
    Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 TATTGTGAAAAAAGAAAAAAG 525
Db 260 TATTGTGAAAAAAGAAAAAAG 283

RESULT 11
ABL70224/c
ID ABL70224 standard; DNA; 7752 BP.
XX
AC ABL70224;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#57.
XX
XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
XX WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPTG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-154758/20.
XX
```

PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling.
XX
PS Claim 1; SEQ ID NO 114; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
XX European Patent Office
XX
SQ Sequence 7752 BP; 1847 A; 113 C; 1696 G; 4096 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 7752;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3349 AAAAAATATATAATAAAAAATAATA 3372
Db 4466 AAAAAATATATAATAAAAAATAATA 4443
|||||

RESULT 12
ABL33535
ID ABL33535 standard; DNA; 8964 BP.
XX
AC ABL33535;
XX
XX 26-MAR-2002 (first entry)
DT
XX Human immune system associated gene SEQ ID NO: 1508.
DE
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
PN
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP007537.
PF
XX
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
PR
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

XX Claim 1; SEQ ID NO 1508; 32pp + Sequence Listing; German.
PS
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAAATATATAATAAAAAATAATA 3373
Db 7648 AAAAAATATATAATAAAAAATAATA 7671
|||||

RESULT 13
ABK31373
ID ABK31373 standard; DNA; 8964 BP.
XX
AC ABK31373;
XX
XX 23-APR-2002 (first entry)
DT
XX Signal transduction associated gene modified complementary DNA #108.
DE
XX Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
XX Homo sapiens.
OS
XX Synthetic.
OS
XX WO200200926-A2.
PN
XX 03-JAN-2002.
PD
XX
XX 29-JUN-2001; 2001WO-EP007472.
PF
XX
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
PR
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-147896/19.
DR
XX
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 216; 24pp; English.
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are

CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAATATATATAATAAAATAATAA 3373
| | | | | | | | | | | | | | | | | |
DB 7648 AAAAATATATATAATAAAATAATAA 7671

RESULT 14
ABL70582
ID ABL70582 standard; DNA; 8964 BP.
XX
AC ABL70582;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence complementary to#236.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPTG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX

XX Claim 1; SEQ ID NO 472; 24pp + Sequence Listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX

SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;

Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3350 AAAAATATATATAATAAAATAATAA 3373
| | | | | | | | | | | | | | | | | |
DB 7648 AAAAATATATATAATAAAATAATAA 7671

RESULT 15
AAS61277
ID AAS61277 standard; DNA; 8964 BP.
XX
AC AAS61277;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #232.
XX

XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX

XX Homo sapiens.
XX
XX WO200177375-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP003968.
XX
XX 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPTG-) EPIGENOMICS AG.
XX

XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-017470/02.
XX

XX New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX

XX Disclosure; SEQ ID NO 238; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 8964 BP; 2764 A; 89 C; 1789 G; 4322 T; 0 U; 0 Other;
Query Match 0.5%; Score 24; DB 6; Length 8964;
Best Local Similarity 100.0%; Pred. Nc. 31;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3350 AAAAATATATATAAAATATATA 3373
Db 7648 AAAAATATATATAAAATATATA 7671

Search completed: July 31, 2004, 11:15:39
Job time : 1574 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 10:32:54 ; Search time 285 Seconds
(without alignments)
8592.976 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatccacagtgataaat.....tttcatgggttcagtgcca 4413

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgm2_6/ptodata/2/ina/ECTUS_COMB.seq.*
6: /cgm2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	23	0.5	367	4 US-09-219-983A-13
2	22	0.5	396	4 US-09-640-173-139
3	22	0.5	396	4 US-09-713-550-139
4	22	0.5	449	4 US-09-621-976-18967
5	22	0.5	1790	4 US-09-866-028-31
6	22	0.5	1866	4 US-09-224-048A-1
7	22	0.5	1664976	4 US-08-916-421B-1
8	21	0.5	35	1 US-08-126-594-20
9	21	0.5	35	1 US-08-465-811A-20
10	21	0.5	35	2 US-08-619-542B-20
11	21	0.5	326	4 US-09-629-645A-17
12	21	0.5	454	4 US-09-621-976-9423
13	21	0.5	780	4 US-09-328-475C-145
14	21	0.5	1198	3 US-08-849-751-3
15	21	0.5	1198	4 US-09-478-816-3
16	21	0.5	1267	1 US-08-140-797-1
17	21	0.5	1267	1 US-08-486-670A-1
18	21	0.5	1292	4 US-09-620-312D-868
19	21	0.5	1415	1 US-08-236-918A-7
20	21	0.5	1415	4 US-09-150-864A-7
21	21	0.5	1586	4 US-09-178-093B-3
22	21	0.5	1601	4 US-09-220-133-182
23	21	0.5	1688	4 US-08-439-814-2
24	21	0.5	1716	2 US-08-954-333-9
25	21	0.5	1779	3 US-09-323-427-1
26	21	0.5	1779	3 US-09-323-427-2
27	21	0.5	1779	4 US-09-812-642-1

```

c 28 21 0.5 1779 4 US-09-812-642-2 Sequence 2, Appli
29 21 0.5 1786 1 US-07-920-430-19 Sequence 19, Appli
30 21 0.5 1786 1 US-08-066-299-9 Sequence 9, Appli
31 21 0.5 1786 1 US-08-265-047-1 Sequence 1, Appli
32 21 0.5 1786 1 US-08-251-464-19 Sequence 19, Appli
33 21 0.5 1786 4 US-08-926-522-20 Sequence 20, Appli
34 21 0.5 1786 5 PCT-US92-01364-19 Sequence 19, Appli
35 21 0.5 2050 4 US-09-688-188B-28 Sequence 28, Appli
36 21 0.5 2050 4 US-09-291-417D-28 Sequence 28, Appli
37 21 0.5 2090 2 US-08-439-814-1 Sequence 1, Appli
38 21 0.5 2274 4 US-09-220-132-188 Sequence 188, App
39 21 0.5 2475 4 US-09-220-132-179 Sequence 179, App
40 21 0.5 2638 4 US-09-228-986-8 Sequence 8, Appli
41 21 0.5 2720 2 US-08-007-107-3 Sequence 3, Appli
42 21 0.5 2738 4 US-09-554-726A-9 Sequence 9, Appli
43 21 0.5 3282 4 US-09-489-847-52 Sequence 52, Appli
44 21 0.5 3347 4 US-09-702-705-318 Sequence 318, App
45 21 0.5 3347 4 US-09-736-457-318 Sequence 318, App

```

ALIGNMENTS

```

RESULT 1
US-09-219-983A-13
; Sequence 13, Application US/09219983A
; Patent No. 6380159
; GENERAL INFORMATION:
; APPLICANT: Wolfner, Mariana
; APPLICANT: Lung, Oliver
; APPLICANT: Tram, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/09/219,983A
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/071,315
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 367
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-219-983A-13

```

```

Query Match 0.5%; Score 23; DB 4; Length 367;
Best Local Similarity 100.0%; Pred.No. 6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 502 TATTGTGAAAAA524
      |||||
Db 336 TATTGTGAAAAA358
      |||||

```

```

RESULT 2
US-09-640-173-139/c
; Sequence 139, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien

```


; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-139

Query Match 0.5%; Score 22; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 TTGTGAAAAAAG 525
Db 27 TTGTGAAAAAAG 6

RESULT 3

US-09-713-550-139/c
; Sequence 139, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-139

Query Match 0.5%; Score 22; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 TTGTGAAAAAAG 525
Db 27 TTGTGAAAAAAG 6

RESULT 4

US-09-621-976-18967
; Sequence 18967, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18967
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18967

Query Match 0.5%; Score 22; DB 4; Length 449;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 TTGTGAAAAAAG 525

Db 428 TTGTGAAAAAAG 449

RESULT 5

US-09-866-028-31
; Sequence 31, Application US/09866028
; Patent No. 6642360
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavini, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tamas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/866,028
; CURRENT FILING DATE: 2001-05-25
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 31
; LENGTH: 1790
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-866-028-31

Query Match 0.5%; Score 22; DB 4; Length 1790;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 503 ATTGTGAAAAAAG 524
Db 1721 ATTGTGAAAAAAG 1742

RESULT 6

US-09-224-048A-1
; Sequence 1, Application US/09224048A
; Patent No. 6387366
; GENERAL INFORMATION:
; APPLICANT: Hurwitz, David R.
; APPLICANT: Cherington, Van
; APPLICANT: Galanopoulos, Theofanis
; APPLICANT: Levine, Peter H.
; APPLICANT: Greenberger, Joel S.
; TITLE OF INVENTION: METHOD FOR REDUCING ADVERSE SIDE EFFECTS ASSOCIATED
; TITLE OF INVENTION: WITH BONE MARROW CELL TRANSPLANTATION
; FILE REFERENCE: 07787/007001
; CURRENT APPLICATION NUMBER: US/09/224,048A
; CURRENT FILING DATE: 1998-12-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Canis familiaris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111)...(986)
US-09-224-048A-1

```
Query Match          0.5%; Score 22; DB 4; Length 1866;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 ATTGTGAAAAAAAAAAAAAAAAAA 524
Db 1834 ATTGTGAAAAAAAAAAAAAAAAAA 1855

RESULT 7
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916.421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
```

```
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1096846)..(1096846)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1119881)..(1119881)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1130881)..(1130881)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1310988)..(1310988)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1313224)..(1313224)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1349473)..(1349473)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1349491)..(1349491)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1470091)..(1470091)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1569020)..(1569020)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1602912)..(1602912)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1603734)..(1603734)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1637998)..(1637998)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ NAME/KEY: misc feature
;/ LOCATION: (1664855)..(1664855)
;/ OTHER INFORMATION: n equals a, t, c, or g
;/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match 0.5%; Score 22; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1353 TATTGAAGATGTTTATATATAT 1374
Db 1283595 TATTGAAGATGTTTATATATAT 1283574
```

```
RESULT 8
US-08-126-594-20/c
; Sequence 20, Application US/08126594
; Patent No. 5482845
; GENERAL INFORMATION:
; APPLICANT: Soares, M. Bento
; APPLICANT: Efstratiadis, Argiris
; TITLE OF INVENTION: METHOD FOR CONSTRUCTION OF NORMALIZED
; TITLE OF INVENTION: CDNA LIBRARIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, c/o Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/126,594
;/ FILING DATE:
;/ CLASSIFICATION: 435
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: White, John P.
;/ REGISTRATION NUMBER: 28,678
;/ REFERENCE/DOCKET NUMBER: 42840/JPW/AKC
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (212) 977-9550
;/ TELEFAX: (212) 664-0525
;/ TELEX: 422523 COOP UI
;/ INFORMATION FOR SEQ ID NO: 20:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 35 base pairs
;/ TYPE: nucleic acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ HYPOTHETICAL: NO
;/ ANTI-SENSE: NO
;/ US-08-126-594-20
```

```
Query Match 0.5%; Score 21; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 509 AAAAAAAAAAAAAAAAAAGTGT 529
Db 34 AAAAAAAAAAAAAAAAAAGTGT 14
```

```
RESULT 9
US-08-465-811A-20/c
; Sequence 20, Application US/08465811A
; Patent No. 5637685
; GENERAL INFORMATION:
; APPLICANT: Soares, M. Bento
; APPLICANT: Efstratiadis, Argiris
; TITLE OF INVENTION: METHOD FOR CONSTRUCTION OF
; TITLE OF INVENTION: NORMALIZED CDNA LIBRARIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White, c/o Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,811A
; FILING DATE: June 6, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
;/ REFERENCE/DOCKET NUMBER: 42840/JPW/AKC
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (212) 977-9550
;/ TELEX: 422523 COOP UI
;/ INFORMATION FOR SEQ ID NO: 20:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 35 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
```

;
: TITLE OF INVENTION: GENES AND GENE EXPRESSION PRODUCTS THAT ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER

;; FILE REFERENCE: 1532.002/200130.463
;; CURRENT APPLICATION NUMBER: US/09/328,475C
;; CURRENT FILING DATE: 1999-06-09
;; NUMBER OF SEQ ID NOS: 341
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 145
;; LENGTH: 780
;; TYPE: DNA
;; ORGANISM: Homo Sapien
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(780)
;; OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-145

Query Match 0.5%; Score 21; DB 4; Length 780;
Best local similarity 100.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 TATTGTGAAAAAAAAAAAAA 522
Db 314 TATTGTGAAAAAAAAAAAAA 334

RESULT 14

US-08-849-751-3
; Sequence 3, Application US/08849751
; Patent No. 6190890
; GENERAL INFORMATION:
; APPLICANT: VAN DEN BROECK, HENRIETTE C.
; APPLICANT: DE GRAAFF, LEENDERT H.
; APPLICANT: VISSER, JACOB
; APPLICANT: VAN OUYEN, ALBERT J.J.
; TITLE OF INVENTION: FUNGAL CELLULASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,751
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/04511
; FILING DATE: 14-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24615-20094.00
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1198 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Aspergillus niger

;; STRAIN: N400
;; INDIVIDUAL ISOLATE: CBS120.49
;; FEATURE:
;; NAME/KEY: Coding Sequence
;; LOCATION: 32...1024
;; OTHER INFORMATION: product="Cellulase"
US-08-849-751-3

Query Match 0.5%; Score 21; DB 3; Length 1198;
Best local similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 TTGTGAAAAAAAAAAAAA 524
Db 1168 TTGTGAAAAAAAAAAAAA 1188

RESULT 15

US-09-478-816-3
; Sequence 3, Application US/09478816
; Patent No. 6306635
; GENERAL INFORMATION:
; APPLICANT: VAN DEN BROECK, HENRIETTE C.
; APPLICANT: DE GRAAFF, LEENDERT H.
; APPLICANT: VISSER, JACOB
; APPLICANT: VAN OUYEN, ALBERT J.J.
; TITLE OF INVENTION: FUNGAL CELLULASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/478,816
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,751
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: PCT/EP96/04511
; FILING DATE: 14-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24615-20094.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1198 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; STRAIN: N400
; INDIVIDUAL ISOLATE: CBS120.49
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 32...1024

```
; OTHER INFORMATION: product="Cellulase"
US-09-478-816-3
Query Match      0.5%; Score 21; DB 4; Length 1198;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      504 TTGTGAAAAAAAAAAAAAAAAA 524
      |||||
Db      1168 TTGTGAAAAAAAAAAAAAAAAA 1188

Search completed: July 31, 2004, 19:48:11
Job time : 291 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 15:50:59 ; Search time 1818 Seconds
(without alignments)
11901.847 Million cell updates/sec

Title: US-10-027-880-3
Perfect score: 4413
Sequence: 1 aagatcacagtgtaataa.....tttcattgggttaagatcca 4413

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3222919 seqs, 2451570024 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4413	100.0	4413	15	US-10-027-880-3
2	2154	48.8	4309	15	US-10-027-880-4
3	2149	48.7	2149	15	US-10-027-880-1
4	2149	48.7	4280	15	US-10-027-880-2
5	26	0.6	15046	15	US-10-311-455-2073
6	25	0.6	68140	13	US-10-087-132-1681
7	24	0.5	388	10	US-09-918-995-8408
8	24	0.5	440	13	US-10-424-599-58606
9	24	0.5	2211	16	US-10-310-154-361
10	24	0.5	8964	13	US-10-221-613-238
11	24	0.5	8964	15	US-10-311-455-1508
12	24	0.5	126266	13	US-10-087-132-1132
13	24	0.5	132762	10	US-09-954-556-17
14	23	0.5	326	10	US-09-803-719-1086

15	23	0.5	367	15	US-10-114-774-13	Sequence 13, Appl
16	23	0.5	390	16	US-10-388-934-308	Sequence 308, App
17	23	0.5	458	9	US-09-764-877-889	Sequence 889, App
18	23	0.5	458	16	US-10-242-515-889	Sequence 889, App
19	23	0.5	613	16	US-10-388-934-310	Sequence 310, App
20	23	0.5	668	13	US-10-027-632-7907	Sequence 7907, App
21	23	0.5	668	16	US-10-027-632-7907	Sequence 7907, App
22	23	0.5	678	13	US-10-424-599-102431	Sequence 102431, App
23	23	0.5	976	9	US-09-764-877-3761	Sequence 3761, App
24	23	0.5	976	16	US-10-242-515-3761	Sequence 3761, App
25	23	0.5	1004	13	US-10-424-599-126393	Sequence 126393, App
26	23	0.5	1734	17	US-10-437-963-29229	Sequence 29229, App
27	23	0.5	1782	13	US-10-424-599-137972	Sequence 137972, App
28	23	0.5	1913	13	US-10-424-599-30424	Sequence 30424, App
29	23	0.5	2532	16	US-10-104-047-326	Sequence 326, App
30	23	0.5	4634	11	US-09-244-805-58	Sequence 58, Appl
31	23	0.5	4634	11	US-09-245-277-58	Sequence 58, Appl
32	23	0.5	4654	13	US-10-221-714-508	Sequence 508, App
33	23	0.5	4654	15	US-10-311-455-2196	Sequence 2196, App
34	23	0.5	6311	15	US-10-311-455-1936	Sequence 1936, App
35	23	0.5	133632	13	US-10-087-192-1810	Sequence 1810, App
36	23	0.5	151858	17	US-10-322-281-653	Sequence 653, App
37	23	0.5	203654	9	US-09-820-905-3	Sequence 3, Appli
38	23	0.5	227246	17	US-10-322-281-314	Sequence 314, App
39	22	0.5	182	9	US-09-867-701-30	Sequence 30, Appl
40	22	0.5	256	9	US-09-867-701-3003	Sequence 3003, App
41	22	0.5	294	10	US-09-918-995-3607	Sequence 3607, App
42	22	0.5	338	13	US-10-424-599-139473	Sequence 139473, App
43	22	0.5	362	15	US-10-198-846-12626	Sequence 12626, App
44	22	0.5	391	15	US-10-058-053A-175	Sequence 175, App
45	22	0.5	396	9	US-09-825-294-139	Sequence 139, App

ALIGNMENTS

RESULT 1

US-10-027-880-3
; Sequence 3, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP:00305
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4413
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-027-880-3

Query Match	100.0%	Score 4413;	DB 15;	Length 4413;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 4413;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AAGATCCACAGTGAATAATAAGAACGGATTCGGTGATATTCGCAACTATATAATGAA	60	
Db	1	AAGATCCACAGTGAATAATAAGAACGGATTCGGTGATATTCGCAACTATATAATGAA	60	
QY	61	ATTGAATCTCTGAGTTCATCGCTTTGGTATCAAGATCGAATCTCTAAAAACATATATCTTA	120	
Db	61	ATTGAATCTCTGAGTTCATCGCTTTGGTATCAAGATCGAATCTCTAAAAACATATATCTTA	120	
QY	121	TAAATAATCTCGAGTGAATTCGATATTTGTAAGCAATCTTGAAAGATATAGTGGG	180	
Db	121	TAAATAATCTCGAGTGAATTCGATATTTGTAAGCAATCTTGAAAGATATAGTGGG	180	

; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS;
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; FILE REFERENCE: CHEP:003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pbin19 vector
; OTHER INFORMATION: Insert
US-10-027-880-4

Query Match 48.8%; Score 2154; DB 15; Length 4309;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 131 CTGCAGTCGAATTCGTGATATATGTAAGCAATCTGAAAAGAAATAGTGGGATATATAAAC 190
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
25 CTGCAGTCGAATTCGTGATATATGTAAGCAATCTGAAAAGAAATAGTGGGATATATAAAC 84
Qy 191 AACGGCGAAGTACAAAGTCTACCTTTTGGCATGGAACCAAGTTTTTGGATTTAC 250
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
85 AACGGCGAAGTACAAAGTCTACCTTTTGGCATGGAACCAAGTTTTTGGATTTAC 144
Qy 251 TTGTGAATTCCTGAATCTTCATTTCTGAATGATATTTACATTTTATCAAAAAAAA 310
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
145 TTGTGAATTCCTGAATCTTCATTTCTGAATGATATTTACATTTTATCAAAAAAAA 204
Qy 311 GTACAAGTCTACCAAGCACAGGATTAACAACACTGTGTGCTCAAAATGCTAATTTAAAG 370
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
205 GTACAAGTCTACCAAGCACAGGATTAACAACACTGTGTGCTCAAAATGCTAATTTAAAG 264
Qy 371 CCTAATCTATGATTTCCCTTTTCTCAGGATATATCTGATATGATGACCCATTT 430
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
265 CCTAATCTATGATTTCCCTTTTCTCAGGATATATCTGATATGATGACCCATTT 324
Qy 431 GTTTGTCAATTAACCTTCCCACTATACATCAGATATCTCAAGTCGAATACATATCCAT 490
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
325 GTTTGTCAATTAACCTTCCCACTATACATCAGATATCTCAAGTCGAATACATATCCAT 384
Qy 491 AAGAAGTGGTATATTTGTGAAAAAAGTGGTATATCTGATATATACAATAACCA 550
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
385 AAGAAGTGGTATATTTGTGAAAAAAGTGGTATATCTGATATATACAATAACCA 444
Qy 551 CGGTCTCGAATTCCTCAACATTTCTAGGAGAAATGGAGTGCTCTTTGGTTTTATT 610
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
445 CGGTCTCGAATTCCTCAACATTTCTAGGAGAAATGGAGTGCTCTTTGGTTTTATT 504
Qy 611 TTATTTCTAATAACATCTATATATTTTAAACCTTCGATGCTCGCTTAAATTTGCAAT 670
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
505 TTATTTCTAATAACATCTATATATTTTAAACCTTCGATGCTCGCTTAAATTTGCAAT 564
Qy 671 GTGCTTAAATTTCTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATG 730
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
565 GTGCTTAAATTTCTAATCATAAATCGTAAAGAAATTCGTGGAAGCCACAGGACATG 624
Qy 731 CATAGGCGAGTAGTTACCTTTAAACCATCAAAAATATATTAATAGAAAAGAACTTC 790
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
625 CATAGGCGAGTAGTTACCTTTAAACCATCAAAAATATATTAATAGAAAAGAACTTC 684
Qy 791 CTAAAGAAACAAATTTAATAGTGAATAAAAAAGATAAGAGGTAGGACAGAAACG 850
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
685 CTAAAGAAACAAATTTAATAGTGAATAAAAAAGATAAGAGGTAGGACAGAAACG 744
Qy 851 TATGGCGGAGCTCGTAAACAGGAGCTCCCGACCACTCGGAGACGGCGAGACGCTGAC 910
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 745 TATGGCGGAGCTCGTAAACAAGGACGTCCTCCGACCACTCGGAGACGGCGAGACGCTGAC 804
Qy 911 TGATTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 970
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
805 TGATTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 864
Qy 971 TATCCAAATTTGTTGCGCTATATATATATAAAGATCCCTCTTTGCTTTGCTTTGCTTTA 1030
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
865 TATCCAAATTTGTTGCGCTATATATATATAAAGATCCCTCTTTGCTTTGCTTTGCTTTA 924
Qy 1031 TTGCGAT 1090
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
925 TTGCGAT 984
Qy 1091 TACTGACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAAAATGACGCCAAAAATTTG 1150
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
985 TACTGACCTCAGTCCCTAGTTAGCTATATGACATATGTGAAAAATGACGCCAAAAATTTG 1044
Qy 1151 AAGAGTTCCCTCTTCTGCAACTAACTCTTACTCTTACTCTTACTCTTACTCTTACTCTTACT 1210
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1045 AAGAGTTCCCTCTTCTGCAACTAACTCTTACTCTTACTCTTACTCTTACTCTTACTCTTACT 1104
Qy 1211 AIGTGTGCACCTCTCGTAAATTAATATGACAGTGCACCTAGATATAAAAAAATGATAGACA 1270
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1105 AIGTGTGCACCTCTCGTAAATTAATATGACAGTGCACCTAGATATAAAAAAATGATAGACA 1164
Qy 1271 TTAGTTTAAACCTGGAATGTTATTTGAACTCTTTTGGATTAACGTGGAATTTGTTATGGA 1330
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1165 TTAGTTTAAACCTGGAATGTTATTTGAACTCTTTTGGATTAACGTGGAATTTGTTATGGA 1224
Qy 1331 TTAATATTTCAAGATAGCA 1390
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1225 TTAATATTTCAAGATAGCA 1284
Qy 1391 GAAAAATATGATGATGATGTTGCTTTTCTAGTACTCTTTTGTGCGTAGTCTCTTT 1450
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1285 GAAAAATATGATGATGATGTTGCTTTTCTAGTACTCTTTTGTGCGTAGTCTCTTT 1344
Qy 1451 CTGCTCATCTCTCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCAT 1510
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1345 CTGCTCATCTCTCTATGAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATTCAT 1404
Qy 1511 AGTCTTTTTCGTAAACCAACAGTCTATGATAGCTATCGTCATCATATCTCTTTGCAACAA 1570
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1405 AGTCTTTTTCGTAAACCAACAGTCTATGATAGCTATCGTCATCATATCTCTTTGCAACAA 1464
Qy 1571 CAAAAAAGATCGTTTTTGTAAAAATTTAGTAGGGCAGTAAAGTCGTCAATTTGTTGCTCGT 1630
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1465 CAAAAAAGATCGTTTTTGTAAAAATTTAGTAGGGCAGTAAAGTCGTCAATTTGTTGCTCGT 1524
Qy 1631 CGAAATCTAGCGTTCTGTCTATCCAAATAGTTTGTGTTGATTCGAGCTTCGAAGTATATA 1690
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1525 CGAAATCTAGCGTTCTGTCTATCCAAATAGTTTGTGTTGATTCGAGCTTCGAAGTATATA 1584
Qy 1691 ATCTTTTGTAGATGGGTCAAGAGATTTCTAACTCGTATACGAGTGTATCCATATATAT 1750
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1585 ATCTTTTGTAGATGGGTCAAGAGATTTCTAACTCGTATACGAGTGTATCCATATATAT 1644
Qy 1751 TCTAACATATACGTTCTTTGTTTTTGGTAGGCTCTCGGCTCTTTTGTAGACACCCCTTGCTA 1810
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1645 TCTAACATATACGTTCTTTGTTTTTGGTAGGCTCTCGGCTCTTTTGTAGACACCCCTTGCTA 1704
Qy 1811 AIGTTTTGTGACCTTAGACAATCCATATAGCTTACGTGAGTGAAGTTGACCAAAA 1870
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1705 AIGTTTTGTGACCTTAGACAATCCATATAGCTTACGTGAGTGAAGTTGACCAAAA 1764
Qy 1871 TGGTCAAAATATATATTTAAATTTGGCCACAAAACAAATTTTACAAACAAATTCACAAA 1930
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1765 TGGTCAAAATATATATTTAAATTTGGCCACAAAACAAATTTTACAAACAAATTCACAAA 1824
Qy 1931 CATGCAATCGTTCAAAATTTTATTTTCAATGCGGTTATTTGTTCAATTTGTAATATCTG 1990
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1825 CATGCAATCGTTCAAAATTTTATTTTCAATGCGGTTATTTGTTCAATTTGTAATATCTG 1884

QY	1991	TTTAACTCAGTACGAAATTTTAAATTTTCAAAGAAGAAACATTTTGTATATAAATAACA	2050
Db	1885	TTTAACTCAGTACGAAATTTTAAATTTTCAAAGAAGAAACATTTTGTATATAAATAACA	1944
QY	2051	TTTTATGGAAACCAACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCTTTGTGAA	2110
Db	1945	TTTTATGGAAACCAACCGGTTAAGCTCGATGATTTTGAGTTTTAGTTTTGTCTTTGTGAA	2004
QY	2111	ATCATTAACGACCTACATATTTGATCCCTCATTTACTCTTAATAATTAGGAATCAAAACATGATG	2170
Db	2005	ATCATTAACGACCTACATATTTGATCCCTCATTTACTCTTAATAATTAGGAATCAAAACATGATG	2064
QY	2171	ATTAAGTTTCAACAAAGACGGTCTTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG	2230
Db	2065	ATTAAGTTTCAACAAAGACGGTCTTTATGGCTATTAAGAGTCAGACGCAAGGATGACCGGG	2124
QY	2231	GTCAATTAAGACGTCCTTATATTCACCACTTACTCCACCTAATTGCTAAATTAATCAG	2284
Db	2125	GTCAATTAAGACGTCCTTATATTCACCACTTACTCCACCTAATTGCTAAATTAATCAG	2178

RESULT 3
 US-10-027-880-1
 ; Sequence 1, Application US/10027880
 ; Publication No. US20030106105A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HOFFMAN, BEATE
 ; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
 ; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
 ; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
 ; FILE REFERENCE: CHEP.003US
 ; CURRENT APPLICATION NUMBER: US/10/027,880
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: PCT/FR00/01768
 ; PRIOR FILING DATE: 2000-06-23
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 2149
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-10-027-880-1

	Query Match	48.7%	Score 2149;	DB 15;	Length 2149;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	136	GTGCAATTGTGATATATTGTAAAGCAATCTGAAAAGAAATAGTGGGATATATAAACAACCG	195		
Db	1	GTGCAATTGTGATATATTGTAAAGCAATCTGAAAAGAAATAGTGGGATATATAAACAACCG	60		
Qy	196	GCAGAACTACAAGTCTACCTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT	255		
Db	61	GCAGAACTACAAGTCTACCTTTTTTGGCATGGAACCATGTTTTTAGGATTTACTTTGT	120		
Qy	256	AATTCTCTGAATCTTTCTTCTTCTTCAATGTATTTTACATTTTTTATCAAAAAAAGTACA	315		
Db	121	AATTCTCTGAATCTTTCTTCTTCTTCAATGTATTTTACATTTTTTATCAAAAAAAGTACA	180		
Qy	316	AGTTCTACCAAGCACAGGAGTTAAACAACCTTGTTGTGCAATGCTAATTTTAAAGCCTAA	375		
Db	181	AGTTCTACCAAGCACAGGAGTTAAACAACCTTGTTGTGCAATGCTAATTTTAAAGCCTAA	240		
Qy	376	TCATTATGAATTCCTCTTTCTTTCAGCATATATACGTATTCATATGCACCCATTTGTTTG	435		
Db	241	TCATTATGAATTCCTCTTTCTTTCAGCATATATACGTATTCATATGCACCCATTTGTTTG	300		
Qy	436	TCATTAACTTCCCACCTCTATACATCAGTATCTCAAAGTCGAATTAACAATATCCATAAGAA	495		
Db	301	TCATTAACTTCCCACCTCTATACATCAGTATCTCAAAGTCGAATTAACAATATCCATAAGAA	360		
Qy	496	GTGGTATATTGTGAAAAAAGGTTGTATCTGTGTATATACAATACCACCGCT	555		

[illegible]

Db 1441 AAGATCGTTTGTAAATTTAGTAGGCACCTAAGTCGTCAATTTGTTGCTCCTCGAAA 1500
Qy 1636 TCTAGCGTTCTGTCAATCCAAATAAGTTGTTGATTCGAGCTTCCAAGATTATAATCTT 1695
Db 1501 TCTAGCGTTCTGTCAATCCAAATAAGTTGTTGATTCGAGCTTCCAAGATTATAATCTT 1560
Qy 1696 TTTTAGATGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATTTCTAA 1755
Db 1561 TTTTAGATGGTCAATGAAGATTTCTAACTTCGTATACGAGTGTATCCATATTTCTAA 1620
Qy 1756 CATATAGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1815
Db 1621 CATATAGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1680
Qy 1816 TTGTTGACCTTAGCAATCCATATACGTTTACGTTGAGTGTGAGTGTGAGTGTGAGTGTG 1875
Db 1681 TTGTTGACCTTAGCAATCCATATACGTTTACGTTGAGTGTGAGTGTGAGTGTGAGTGTG 1740
Qy 1876 CAATATATATTTAAATTTGGCCCAAAACAAACATTTTACAAACAAATTTCAACAAATGC 1935
Db 1741 CAATATATATTTAAATTTGGCCCAAAACAAACATTTTACAAACAAATTTCAACAAATGC 1800
Qy 1936 ATCGTTTCAAAATTTTATTTTCAATGCGGTTTATTTGTTGTTGTTGTTGTTGTTGTT 1995
Db 1801 ATCGTTTCAAAATTTTATTTTCAATGCGGTTTATTTGTTGTTGTTGTTGTTGTTGTT 1860
Qy 1996 CTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTGTATATATTAATTAATTAATTA 2055
Db 1861 CTCACTGACGAATTTTAAATTTTCAAGAGAACATTTTGTATATATTAATTAATTAATTA 1920
Qy 2056 TGAACACCGGTTAAGCTCGATGATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 2115
Db 1921 TGAACACCGGTTAAGCTCGATGATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1980
Qy 2116 TAAGACCTCAATTTGATCCCTCATTTACTTTAATTAATTAAGGAATCAACATGATGATTA 2175
Db 1981 TAAGACCTCAATTTGATCCCTCATTTACTTTAATTAATTAAGGAATCAACATGATGATTA 2040
Qy 2176 GTTCACCAAGAGCTCTTATGCTATTAAGCTATTAAGCTATTAAGCTATTAAGCTATTAAGCT 2235
Db 2041 GTTCACCAAGAGCTCTTATGCTATTAAGCTATTAAGCTATTAAGCTATTAAGCTATTAAGCT 2100
Qy 2236 TAAGACCTCAATTTAATTTCAACCATTTACTCCATTAATGCTAATTAATTAATTAATTA 2284
Db 2101 TAAGACCTCAATTTAATTTCAACCATTTACTCCATTAATGCTAATTAATTAATTAATTA 2149

RESULT 4

US-10-027-880-2
; Sequence 2, Application US/10027880
; Publication No. US20030106105A1
; GENERAL INFORMATION:
; APPLICANT: HOFFMAN, BEATE
; TITLE OF INVENTION: PROMOTER EXPRESSED SPECIFICALLY IN PLANT ROOT CELLS,
; TITLE OF INVENTION: VECTORS AND RECOMBINANT HOST CELLS CONTAINING SUCH A
; TITLE OF INVENTION: PROMOTER AND TRANSGENIC PLANTS OBTAINED
; FILE REFERENCE: CHEP-003US
; CURRENT APPLICATION NUMBER: US/10/027,880
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: PCT/FR00/01768
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4280
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Construction
; OTHER INFORMATION: promoter + sequence coding for the gus gene

US-10-027-880-2

Query Match 48.7%; Score 2149; DB 15; Length 4280;

Best Local Similarity 100.0%; Pred. No. 0; Matches 2149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 136 GTGAAATGTGATATATTGTAAGCAATCTGAAAGAAATAAGTGGGATATATAAACAACCG 195
Db 1 GTCGAAATGTGATATATTGTAAGCAATCTGAAAGAAATAAGTGGGATATATAAACAACCG 60
Qy 196 GCGAAAGTACAGTTCTACCTTTTGGCATGGAACCAATGTTTGGATGTTTGGATGTTTGGT 255
Db 61 GCGAAAGTACAGTTCTACCTTTTGGCATGGAACCAATGTTTGGATGTTTGGATGTTTGGT 120
Qy 256 AATTCTGGAATCTTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGATACA 315
Db 121 AATTCTGGAATCTTTTCAATTTCTGAAATGATATTTACATTTTATCAAAAAAAGATACA 180
Qy 316 AGTTCTACCAAGACAGGAGTTAAACAACCTGTGTCTCAAAATGCTAATTTAAAGCCATA 375
Db 181 AGTTCTACCAAGACAGGAGTTAAACAACCTGTGTCTCAAAATGCTAATTTAAAGCCATA 240
Qy 376 TCTTATGATTTCCCTTTTCTTACGATATATCTGATATGATATGACCCCATTTGTTG 435
Db 241 TCTTATGATTTCCCTTTTCTTACGATATATCTGATATGATATGACCCCATTTGTTG 300
Qy 436 TCAATTAATTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATAATCCATAAGAA 495
Db 301 TCAATTAATTTCCCACTCTATACATCAGTATCTCAAAGTCGAATAACAATAATCCATAAGAA 360
Qy 496 GTGGTATATTTGTAAGAAAAAAGTGTGTATCTGGTATATACATACACACCGTC 555
Db 361 GTGGTATATTTGTAAGAAAAAAGTGTGTATCTGGTATATACATACACACCGTC 420
Qy 556 TCGAATTTGCTTCAACATTTCTAGGAGAAAAATGAGCTGTCTCTTGGTTTATTTTATTT 615
Db 421 TCGAATTTGCTTCAACATTTCTAGGAGAAAAATGAGCTGTCTCTTGGTTTATTTTATTT 480
Qy 616 CTTAATAACATATCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTTCAATGTGCC 675
Db 481 CTTAATAACATATCTATATTTTAAACACATTCGATGTCTCGCTTAAATTTTCAATGTGCC 540
Qy 676 TAAATTTCTTAATCATATAATCGTAAGAAAAATTCGTGAGAACACACAGGACATGATAG 735
Db 541 TAAATTTCTTAATCATATAATCGTAAGAAAAATTCGTGAGAACACACAGGACATGATAG 600
Qy 736 GGCAGGTAGTTTACCTTTTAAACCATCAAAATATATTAATAAGAAAGAACTTCCATAA 795
Db 601 GGCAGGTAGTTTACCTTTTAAACCATCAAAATATATTAATAAGAAAGAACTTCCATAA 660
Qy 796 AGAACAAATTTATAAAGTGGATAAAAAGATAAGAGGTAGGCAGAGAAACGATATGG 855
Db 661 AGAACAAATTTATAAAGTGGATAAAAAGATAAGAGGTAGGCAGAGAAACGATATGG 720
Qy 856 CCGGACCTGTAACAGAGGACGTCGCCACCTCGGAGACGGGAGACGCTGATGATT 915
Db 721 CCGGACCTGTAACAGAGGACGTCGCCACCTCGGAGACGGGAGACGCTGATGATT 780
Qy 916 TTTTCTTTTCTTTTCTTAAAGAAAGTGTGTTGCTGCTTCAAGGGTCAAAACCATATCC 975
Db 781 TTTTCTTTTCTTTTCTTAAAGAAAGTGTGTTGCTGCTTCAAGGGTCAAAACCATATCC 840
Qy 976 AATTGTTCTGCTTATTTATTAATACTAAAGATCCCTCTTGTCTTGTCTTGTCTTGTCT 1035
Db 841 AATTGTTCTGCTTATTTATTAATACTAAAGATCCCTCTTGTCTTGTCTTGTCTTGTCT 900
Qy 1036 GATATATAATCTAACTTAAATTTAGTTTCTAAATAATATATATGCTCTACCTATGTTCTACTG 1095
Db 901 GATATATAATCTAACTTAAATTTAGTTTCTAAATAATATATATGCTCTACCTATGTTCTACTG 960
Qy 1096 ACCTCAGTCCCTAGTTAGCTATATGACATATGGAATAATGAGCCCAAAATTTGAAGAG 1155
Db 961 ACCTCAGTCCCTAGTTAGCTATATGACATATGGAATAATGAGCCCAAAATTTGAAGAG 1020
Qy 1156 TTCTCTCTCTCTGCACTTAATCTTATCTTATCTANTGAGCTATGTTAAATATGAAATGTT 1215

Db 1021 TTCCTCTCCGCAACTCTTAICTTACTCATGTGACGTATGTTAAATATTTGAATGTT 1080

Qy 1216 GGCACCTCGTATTAAATATGCGAGTTCACCTAGATATAAAAAACATGATAGACATTTAG 1275

Db 1081 GGCACCTCGTATTAAATATGCGAGTTCACCTAGATATAAAAAACATGATAGACATTTAG 1140

Qy 1276 TTTAAACCTGAAATGTTATTGAACTCTTTGGAATACGTGAATGTTGATGGATTAAA 1335

Db 1141 TTTAAACCTGAAATGTTATTGAACTCTTTGGAATACGTGAATGTTGATGGATTAAA 1200

Qy 1336 TTTTGAAGATTTTATATATTGAAGATGTTTATATATATAGATTTATAGACGAAAA 1395

Db 1201 TTTTGAAGATTTTATATATTGAAGATGTTTATATATATAGATTTATAGACGAAAA 1260

Qy 1396 TATTGATGTAGATGTTGTCCTTTTCTGTTAGTACTCTTTTGTGGGTAGTCTCTTCTCT 1455

Db 1261 TATTGATGTAGATGTTGTCCTTTTCTGTTAGTACTCTTTTGTGGGTAGTCTCTTCTCT 1320

Qy 1456 CATCTCCTATGAAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATCATAGTCT 1515

Db 1321 CATCTCCTATGAAGAAAAATCCAAATAGTTTAAAGAAATTTTGTGTAATCATAGTCT 1380

Qy 1516 TTTTCGTAAACACAGTCTATGCTATGCTATCGTCACTATTCCTCTTTCGCAACAAAA 1575

Db 1381 TTTTCGTAAACACAGTCTATGCTATGCTATCGTCACTATTCCTCTTTCGCAACAAAA 1440

Qy 1576 AAGATCGTTTGTAAAAATTTAGTAGGGCACTAAAGTCGTCAATTTGTTGTCCTGCGAAA 1635

Db 1441 AAGATCGTTTGTAAAAATTTAGTAGGGCACTAAAGTCGTCAATTTGTTGTCCTGCGAAA 1500

Qy 1636 TCTAGGCTTCTGTCATCCACAAATAGTTTGTGTTGATTCGAGCTTCCAGATTAATATCTT 1695

Db 1501 TCTAGGCTTCTGTCATCCACAAATAGTTTGTGTTGATTCGAGCTTCCAGATTAATATCTT 1560

Qy 1696 TTTTATAGTGGTCACTGAAGATTTCTAACTTCGTATACGAGTATCCCATATATTTCTAA 1755

Db 1561 TTTTATAGTGGTCACTGAAGATTTCTAACTTCGTATACGAGTATCCCATATATTTCTAA 1620

Qy 1756 CATATACGTCCTGTTTGTAGGCTCTGCGTCTTTTGAGACACCCCTTCTTAATGTT 1815

Db 1621 CATATACGTCCTGTTTGTAGGCTCTGCGTCTTTTGAGACACCCCTTCTTAATGTT 1680

Qy 1816 TTGTTGCACCTTAGCAATCCATATAATGTTTGTGAGTCGAGTCGACCAAAATGGTC 1875

Db 1681 TTGTTGCACCTTAGCAATCCATATAATGTTTGTGAGTCGAGTCGACCAAAATGGTC 1740

Qy 1876 CAATATATATTTAAATTTGGCCACAAACAAATTTACAAACAAATTTCAACAAATGTC 1935

Db 1741 CAATATATATTTAAATTTGGCCACAAACAAATTTACAAACAAATTTCAACAAATGTC 1800

Qy 1936 ATCGTTCCTCAATTTTATTTCAATGCGGTTATTTGTTCAATGTAATATTTCTGTTAA 1995

Db 1801 ATCGTTCCTCAATTTTATTTCAATGCGGTTATTTGTTCAATGTAATATTTCTGTTAA 1860

Qy 1996 CTCACGTAGCAATTTTAAATTTTCAAGAAACAAATTTTGAATATAAATAACATTTTA 2055

Db 1861 CTCACGTAGCAATTTTAAATTTTCAAGAAACAAATTTTGAATATAAATAACATTTTA 1920

Qy 2056 TCGAACCCCGTTAAGCTCGATGATTTTGTAGTTTGTGTTTGTGTTTGTGAATCAT 2115

Db 1921 TCGAACCCCGTTAAGCTCGATGATTTTGTAGTTTGTGTTTGTGTTTGTGAATCAT 1980

Qy 2116 TAAACACCTTACATTTTATCCCTTACTTACTTAAATAGGAATCAACATGATGATTAA 2175

Db 1981 TAAACACCTTACATTTTATCCCTTACTTACTTAAATAGGAATCAACATGATGATTAA 2040

Qy 2176 GTTACCAAGACGTCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATACCGGGTCTAT 2235

Db 2041 GTTACCAAGACGTCTCTTATGGCTATTAAAGAGTCAGACGCAAGGATACCGGGTCTAT 2100

Qy 2236 TAAGAGCTCTTATATCAACCATTTACTCCACTAAATGCTTAATATCATG 2284

Db 2101 TAAGAGCTCTTATATCAACCATTTACTCCACTAAATGCTTAATATCATG 2149

RESULT 5

US-10-311-455-2073/c
; Sequence 2073, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2073
; LENGTH: 15046
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2073

Query Match 0.6%; Score 26; DB 15; Length 15046;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3349 AAAAAATATATATAAAAAATATAAA 3374
Db 9895 AAAAAATATATATAAAAAATATAAA 9870

RESULT 6

US-10-087-192-1681
; Sequence 1681, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1681
; LENGTH: 68140
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) - (68140)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1681

Query Match 0.6%; Score 25; DB 13; Length 68140;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 TGAATAAAAAAATAAAAGTGTAT 531
|||||

Db 47578 TGAAGAAAAAAGAGTGTAT 47602

RESULT 7

US-09-918-995-8408
; Sequence 8408, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8408
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-8408

Query Match 0.5%; Score 24; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 TATTGTGAAAAAAGAGAG 525
|||||
Db 260 TATTGTGAAAAAAGAGAG 283

RESULT 8

US-10-424-599-58606/c
; Sequence 58606, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 58606
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_23933C.1
US-10-424-599-58606

Query Match 0.5%; Score 24; DB 13; Length 440;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 501 ATATTGTGAAAAAAGAGAG 524
|||||
Db 75 ATATTGTGAAAAAAGAGAG 52

RESULT 9

US-10-310-154-361
; Sequence 361, Application US/10310154
; Publication No. US20030233670A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Chomet, Paul S.
; APPLICANT: Adams, Thomas H

; APPLICANT: Ruff, Thomas G.
; APPLICANT: Agarwal, Ameeta K.
; APPLICANT: Ahrens, Jeffrey E.
; APPLICANT: Ball, James A.
; APPLICANT: Banu, G.
; APPLICANT: Bell, Erin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Deng, Molian
; APPLICANT: Dong, Jinzhao
; APPLICANT: Duff, Stephen M.
; APPLICANT: Gallican, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shihshieh
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A.
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Lin, Jie-Yi
; APPLICANT: Liu, Jingdong
; APPLICANT: Lu, Bin
; APPLICANT: Luethy, Michael M.
; APPLICANT: Madson, Linda L.
; APPLICANT: Malloy, Kathleen A.
; APPLICANT: McKiel, Christine L.
; APPLICANT: Miller, Philip W.
; APPLICANT: Padmavathi, Manchikanti
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Start, William G.
; APPLICANT: Tennesen, Dan
; APPLICANT: Vidya, K.R.
; APPLICANT: Wang, Haiyun
; APPLICANT: Xin, Zhaunguo
; APPLICANT: Xu, Nanfei
; APPLICANT: Yang, Chunzhi
; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, Yajuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310,154
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; PRIOR FILING DATE: 2001-12-04
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 361
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)..(1634)
; OTHER INFORMATION:
US-10-310-154-361

Query Match 0.5%; Score 24; DB 16; Length 2211;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 502 TATTGTGAAAAAAGAG 525
|||||
Db 2188 TATTGTGAAAAAAGAG 2211

RESULT 10

US-10-221-613-238
; Sequence 238, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:

```
/
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 238
; LENGTH: 8964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-238

Query Match      0.5%; Score 24; DB 13; Length 8964;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAAATATATAATAAAAAATAATAA 3373
      |||||||
Db 7648 AAAAAATATATAATAAAAAATAATAA 7671

RESULT 11
US-10-311-455-1508
; Sequence 1508, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Expression of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1508
; LENGTH: 8964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1508

Query Match      0.5%; Score 24; DB 15; Length 8964;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAAATATATAATAAAAAATAATAA 3373
      |||||||
Db 7648 AAAAAATATATAATAAAAAATAATAA 7671

; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; PRIOR FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 238
; LENGTH: 8964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-238

Query Match      0.5%; Score 24; DB 13; Length 8964;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAAATATATAATAAAAAATAATAA 3373
      |||||||
Db 7648 AAAAAATATATAATAAAAAATAATAA 7671

RESULT 11
US-10-311-455-1508
; Sequence 1508, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Expression of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1508
; LENGTH: 8964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1508

Query Match      0.5%; Score 24; DB 15; Length 8964;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3350 AAAAAATATATAATAAAAAATAATAA 3373
      |||||||
Db 7648 AAAAAATATATAATAAAAAATAATAA 7671
```

```
RESULT 12
US-10-087-192-1132/c
; Sequence 1132, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: MORRIS, David W.
; APPLICANT: ENGELHARD, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1132
; LENGTH: 126266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(126266)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1132

Query Match      0.5%; Score 24; DB 13; Length 126266;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 AAAAAAAAAAAAAAAAAAGTGGTATA 532
      |||||||
Db 59913 AAAAAAAAAAAAAAAAAAGTGGTATA 59890

RESULT 13
US-09-954-556-17/c
; Sequence 17, Application US/09954556
; Publication No. US20030078219A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Susan M. Freier
; APPLICANT: Scott Cooper
; TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 2 EXPRESSION
; FILE REFERENCE: RTS-0250
; CURRENT APPLICATION NUMBER: US/09/954,556
; CURRENT FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 108
; SEQ ID NO 17
; LENGTH: 132762
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(132762)
; OTHER INFORMATION: n = A,T,C or G
US-09-954-556-17

Query Match      0.5%; Score 24; DB 10; Length 132762;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 AAAAAAAAAAAAAAAAAAGTGGTATA 532
      |||||||
Db 54339 AAAAAAAAAAAAAAAAAAGTGGTATA 54316

RESULT 14
US-09-803-719-1086/c
; Sequence 1086, Application US/09803719
```

```

; Publication No. US20030044783A1
;
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Williams, Jaime
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominiguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radjoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
;
; TITLE OF INVENTION: Human Genes and Gene Products
;
; FILE REFERENCE: 1634.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1086
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/key: misc_feature
; LOCATION: (1)...(326)
;
; OTHER INFORMATION: n = A,T,C or G
US-09-803-719-1086

```

```
Query Match          0.5%; Score 23; DB 10; Length 326;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 509 AAAAAAAAAAAAAAAAAAAGTGGTAT 531
|||
Db 280 AAAAAAAAAAAAAAAAAAAGTGGTAT 258

RESULT 15
US-10-114-774-13
; Sequence 13, Application US/10114774
; Publication No. US20030166533A1
; GENERAL INFORMATION:
; APPLICANT: Wolfner, Mariana
; APPLICANT: Lung, Oliver
; APPLICANT: Tram, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; TITLE OF INVENTION: MELANOGASTER
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/10/114, 774
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/219, 983
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 367
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-114-774-13

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2004, 10:22:14 ; Search time 10330 Seconds
(without alignments)

12757.193 Million cell updates/sec

Title: US-10-027-880-3

Perfect score: 4413

Sequence: 1 aaatcacacagtgaataat.....tttcatgggttaacagatcca 4413

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gsl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	355	8.0	355	29	CC793681 SALK 0173
2	201	4.6	452	28	BH611124 SALK_0299
3	103	2.3	501	9	AV442527 AV442527
4	94	2.1	619	9	AV832242 AV832242

c	5	75	1.7	379	28	BH757746
	6	73	1.7	575	9	AV552576
	7	67	1.5	74	28	BZ595332
	8	55	1.2	29	BX285678	Arabidops
	9	44	1.0	83	29	AL763344
	10	30	0.7	703	28	BZ470713
	11	30	0.7	833	28	BZ457686
	12	29	0.7	874	28	BZ508310
	13	26	0.6	646	28	BH963752
	14	26	0.6	691	28	BH942846
	15	26	0.6	736	28	BH438688
	16	26	0.6	1209	28	CC258283
	17	25	0.6	128	14	CD420389
	18	25	0.6	281	9	AW016688
	19	25	0.6	300	14	CD606616
	20	25	0.6	307	14	CD605249
	21	25	0.6	329	9	AW004306
	22	25	0.6	350	9	AI838645
	23	25	0.6	388	14	CD420251
	24	25	0.6	410	10	AW575585
	25	0.6	418	14	CF273396	EST3085
	26	25	0.6	442	9	AA925170
	27	25	0.6	445	10	AW165189
	28	25	0.6	465	12	BM673424
	29	25	0.6	478	13	BU764603
	30	25	0.6	479	10	AW121179
	31	25	0.6	579	14	CB851263
	32	25	0.6	592	10	AW424360
	33	25	0.6	598	14	CD374151
	34	25	0.6	603	12	BQ003582
	35	25	0.6	619	13	EU238032
	36	25	0.6	637	14	CB588141
	37	25	0.6	639	14	CB823573
	38	25	0.6	681	13	BQ628781
	39	25	0.6	685	13	BQ705470
	40	25	0.6	721	29	CE625066
	41	25	0.6	741	12	BQ008193
	42	25	0.6	758	13	BQ633081
	43	25	0.6	2441	11	BC052222
	44	24	0.5	96	9	AA432834
	45	24	0.5	142	12	BI133442

ALIGNMENTS

RESULT 1
CC793681

LOCUS

DEFINITION

CC793681

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CC793681 355 bp DNA linear GSS 01-JUL-2003
SALK_017399.55.00.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_017399.55.00.x, genomic
survey sequence.

CC793681 GI:32388904

GSS.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 355)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,

Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,

Shinn,P., Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

ORIGIN

```

RESULT 6
AV552576
LOCUS
DEFINITION AV552576 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone R233e03R 5', mRNA sequence.
ACCESSION AV552576
VERSION AV552576
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 575)
REFERENCE
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7 (3), 175-180 (2000)
MEDLINE 20363093
PubMed 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..575
    /organism="Arabidopsis thaliana"
    /mol_type="mRNA"
    /strain="Columbia"
    /db_xref="taxon:3702"
    /clone="R233e03R"
    /tissue_type="roots"
    /clone_lib="Arabidopsis thaliana roots Columbia"
    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 1.7%; Score 73; DB 9; Length 575;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3734 GAGCATGTCGGTTACGAGATCTGCGACGACCTGTCGACATCAGCCAGTCTCGCT 3793
      |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GAGCATGTCGGTTACGAGATCTGCGACGACCTGTCGACATCAGCCAGTCTCGCT 60

QY 3794 TCGTAATGTTACG 3806
      |||||||||||
Db 61 TCGTAATGTTACG 73

RESULT 7
BZ595332
LOCUS
DEFINITION BZ595332 43 40 x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_086892.43.40.x, genomic
            survey sequence.
ACCESSION BZ595332
VERSION BZ595332
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 74)
REFERENCE
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
Shinn, P., Zimmerman, J. and Ecker, J.R.
TITLE A Sequence-indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished (2001)

```

```

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within an annotated intron of Atlg77750.
Class: TDNA tagged.

FEATURES
source
1..74
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
    /strain="Columbia 0"
    /db_xref="taxon:3702"
    /clone="SALK_086892.43.40.x"
    /clone_lib="Arabidopsis thaliana TDNA insertion lines"
    /note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 1.5%; Score 67; DB 28; Length 74;
Best Local Similarity 100.0%; Pred. No. 4.3e-06;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4347 TCCTCTTTAGATTATACATCAATTATGTTATTTCTTCTGCTACAGTTTCATGGGTAC 4406
      |||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TCCTCTTTAGATTATACATCAATTATGTTATTTCTTCTGCTACAGTTTCATGGGTAC 60

QY 4407 GAGTCCA 4413
      |||||
Db 61 GAGTCCA 67

RESULT 8
BX285678/c
LOCUS
DEFINITION BX285678 Arabidopsis thaliana T-DNA flanking sequence GK-384F07-017271,
            genomic survey sequence.
ACCESSION BX285678
VERSION BX285678.1
KEYWORDS GI:28884674
SOURCE GSS.
ORGANISM Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.
AUTHORS and Weissshaar, B.
TITLE A pipeline for automated high-throughput generation of ESTs
(fanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso, M., Strizhov, N., Li, Y. and Weissshaar, B.
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso, M., Strizhov, N., Li, Y. and Weissshaar, B.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2003) Weissshaar B., Max-Planck-Institut fuer
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T32E8. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German

```


Plant Genomics program designated 'GABI'. Information on line availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1. .92

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-384F07-017271"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

ORIGIN

Query Match 1.2%; Score 55; DB 29; Length 92;

Best Local Similarity 100.0%; Pred. No. 0.0028;

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2008 TTTTAAATTTTCAAGAAGACATTTTCATATAAATAACATTTTATGGAACC 2062

Db 92 TTTTAAATTTTCAAGAAGACATTTTCATATAAATAACATTTTATGGAACC 38

RESULT 9

AL763344/c

LOCUS

DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-036B01-015001, genomic survey sequence.

ACCESSION AL763344

VERSION AL763344.1 GI:21512449

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.

and Weissshaar, B.

A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

Unpublished

JOURNAL

REFERENCE

AUTHORS

TITLE

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

for flanking sequence tag based reverse genetics

JOURNAL

REFERENCE

AUTHORS

TITLE

Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone T3288. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1. .83

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-036B01-015001"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

ORIGIN

Query Match 1.0%; Score 44; DB 29; Length 83;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1638 TAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCC 1681

Db 83 TAGCGTTCTGTCATCCACAAATAAGTTGTTTGATTCGAGCTTCC 40

RESULT 10

BZ470713

LOCUS

DEFINITION BZ470713 703 bp DNA linear GSS 13-DEC-2002 BZ470713 BO_1.6.2_KB_tot Brassica oleracea genomic clone BONNA88, genomic survey sequence.

ACCESSION BZ470713

VERSION BZ470713.1 GI:26768062

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM

Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 703)
 Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Other GSSs: BONNA88TF
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

Location/Qualifiers

1. .703

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BONNA88"

/clone_lib="BO_1.6.2_KB_tot"

/note="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared

total DNA inserted into pHO51 using BstXI linkers"

ORIGIN

Query Match 0.7%; Score 30; DB 28; Length 703;

Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 882 GACCACTCGGAGACGGCGAGACGCTGACT 911

Db 660 GACCACTCGGAGACGGCGAGACGCTGACT 689

RESULT 11

BZ457686/c

LOCUS BZ457686 833 bp DNA linear GSS 13-DEC-2002
DEFINITION BONRW80TR BO 1.6 2 KB tot Brassica oleracea genomic clone BONRW80,
genomic survey sequence.
ACCESSION BZ457686
VERSION BZ457686.1 GI:26737311
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 833)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BONRW80TF
Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..833
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONRW80"
/clone_lib="BO 1.6 2 KB tot"
/note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN
Query Match 0.7%; Score 30; DB 28; Length 833;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 882 GACCACTGCGGAGCGGAGCGCTGACT 911
|||||
Db 647 GACCACTGCGGAGCGGAGCGCTGACT 618
|||||
RESULT 12
BZ508310/c
LOCUS BZ508310 874 bp DNA linear GSS 16-DEC-2002
DEFINITION BONHJ80TR BO 1.6 2 KB tot Brassica oleracea genomic clone BONHJ80,
genomic survey sequence.
ACCESSION BZ508310
VERSION BZ508310.1 GI:27031022
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 874)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BONHJ80TF
Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
1..874
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONHJ80"
/clone_lib="BO 1.6 2 KB tot"
/note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN
Query Match 0.7%; Score 29; DB 28; Length 874;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2314 TGATTAGTAGTCTCAGCCACAGGCAA 2342
|||||
Db 301 TGATTAGTAGTCTCAGCCACAGGCAA 273
|||||
RESULT 13
BH963752/c
LOCUS BH963752 646 bp DNA linear GSS 01-OCT-2002
DEFINITION odf98b01.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH963752
VERSION BH963752.1 GI:23444978
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 646)
AUTHORS Delehaanty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odf98 row: b column: 01
Seq primer: -21UPPOT forward
Class: shotgun
High quality sequence start: 64
High quality sequence stop: 551.
FEATURES
source
1..646
Location/Qualifiers
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TO1000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (GSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
ORIGIN
Query Match 0.6%; Score 26; DB 28; Length 646;
Best Local Similarity 100.0%; Pred. No. 4.2e+03;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 870 AAGGACGTCCTCCGACCACTGCGGAGA 895
|||||
Db 515 AAGGACGTCCTCCGACCACTGCGGAGA 490
|||||
RESULT 14

BH942846/c
 LOCUS BH942846 691 bp DNA linear GSS 01-OCT-2002
 DEFINITION odel19f10.b1 B.Oleracea002 Brassica oleracea genomic, genomic survey sequence.
 ACCESSION BH942846
 VERSION BH942846.1 GI:23422906
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 691)
 AUTHORS Delehaanty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash,W., Rabinowicz,P.D. and Wilson,R.K.
 TITLE Whole genome shotgun reads from Brassica oleracea
 JOURNAL Unpublished (2002)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: odel9 row: f column: 10
 Seq primer: -21UPPOT forward
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 551.

FEATURES

source 1..691
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.Oleracea002"

/note="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T01000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 0.6%; Score 26; DB 28; Length 691;
 Best Local Similarity 100.0%; Pred.No. 3.9e+03;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 283 TGATATTACATTTTATCAAAAAA 308
 ||||||||||||||||||||
 Db 610 TGATATTACATTTTATCAAAAAA 585

RESULT 15

BH438688
 LOCUS BH438688 736 bp DNA linear GSS 12-DEC-2001
 DEFINITION BOGGM25TR BOGG Brassica oleracea genomic clone BOGGM25, genomic survey sequence.

ACCESSION BH438688
 VERSION BH438688.1 GI:17624402
 KEYWORDS GSS.
 SOURCE Brassica oleracea

ORGANISM

Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 736)

Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOGGM25TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES

source 1..736
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOGGM25"
 /clone_lib="BOGG"

/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 0.6%; Score 26; DB 28; Length 736;
 Best Local Similarity 100.0%; Pred.No. 3.7e+03;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 283 TGATATTACATTTTATCAAAAAA 308
 ||||||||||||||||||||
 Db 192 TGATATTACATTTTATCAAAAAA 217

Search completed: July 31, 2004, 18:43:15
 Job time : 10338 secs

This Page Blank (uspto)